DOCUMENT RESUME

ED 364 583 TM 020 809

TITLE	Report on Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices. Project 3.1: Alternative Assessments in Schools: The Effects of Alternative Assessments.
INSTITUTION	National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Aug 93
CONTRACT	R117G10027
NOTE	108p.; Papers presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Accountability; Case Studies; Classroom Techniques; *Educational Assessment; Educational Attitudes; Educational Practices; Elementary Education; *Elementary School Mathematics; Knowledge Level; Learning; Literacy; Parent Attitudes; Research Methodology; Standardized Tests; *Student Evaluation; Teacher Attitudes; Teacher Developed Materials; Teaching Methods
IDENTIFIERS	*Alternative Assessment; Assessment Effects; Learner Based Accountability System; *Performance Based Evaluation

ABSTRACT

Four papers follow and document the progress of the Center for Research on Evaluation, Standards, and Student Testing in understanding questions related to alternative assessment based on case studies in three elementary schools. The research questions include consideration of the logistical constraints that must be respected in developing alternative assessments and the features that can be feasibly integrated with instruction. Also considered are changes in teachers' knowledge and beliefs about assessment and instruction resulting from the project and the effect of new assessments on learning. The four papers are: (1) "Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics" (Roberta A. Flexer and Eileen A. Gerstner); (2) "Dilemmas and Issues in Implementing Classroom-Based Assessments for Literacy" (Elfrieda H. Hiebert and Kathryn Davinroy); (3) "Teachers' Ideas and Practices about Assessment and Instruction" (Hilda Borko, Maurene Flory, and Kate Cumbo); and (4) "Parent Opinions about Standardized Tests, Teacher's Information and Performance Assessments" (Lorrie A. Shepard and Carribeth L. Bliem). (Contains 57 references.) (SLD)

ED 364 583

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – August 1993

Project 3.1 Alternative Assessments in Schools
The Effects of Alternative Assessments

Report on Case Study of the Effects of Alternative Assessment in Instruction, Student Learning and Accountability Practices

Lorrie Shepard, Project Director
CRESST/University of Colorado at Boulder

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

TM020809

BEST COPY AVAILABLE

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

INTRODUCTION

Center for Research on Evaluation, Standards, and Student Testing (CRESST)

The current intense interest in alternative forms of assessment is based on a number of assumptions that are as yet untested. In particular, the claim that authentic assessments will improve instruction and student learning is supported only by negative evidence from research on the effects of traditional multiple-choice tests. Because it has been shown that student learning is reduced by teaching to tests of low level skills, it is theorized that teaching to more curricularly defensible tests will improve student learning (Frederiksen & Collins, 1989; Resnick & Resnick, 1992). Studies in Improving Classroom and Local Assessments examine the actual effects of introducing new forms of assessment at the classroom level.

Derived from theoretical arguments about the anticipated effects of authentic assessments and from the framework of past empirical studies that examined the effects of standardized tests (Shepard, 1991), the study examines a number of inter-related research questions:.

1. What logistical constraints must be respected in developing alternative assessments for classroom purposes? What are the features of assessments that can feasibly be integrated with instruction?

2. What changes occur in teachers' knowledge and beliefs about assessment as a result of the project? What changes occur in classroom assessment practices? Are these changes different in writing, reading, and mathematics, or by type of school?

3. What changes occur in teachers' knowledge and beliefs about instruction as a result of the project? What changes occur in instructional practices? Are these changes different in writing, reading, and mathematics, or by type of school?

4. What is the effect of new assessments on student learning? What picture of student learning is suggested by improvements as measured by the new assessments? Are gains in student achievement corroborated by external measures?

5. What is the impact of new assessments on parents' understandings of the curriculum and their children's progress? Are new forms of assessment credible to parents and other "accountability audiences" such as school boards and accountability committees?

The four papers will follow document our progress in understanding these questions, based on case studies in three elementary schools.

TABLE OF CONTENTS

INTRODUCTION v

DILEMMAS AND ISSUES FOR TEACHERS DEVELOPING PERFORMANCE ASSESSMENTS IN MATHEMATICS 1
Roberta J. Flexer and Eileen A. Gerstner

DILEMMAS AND ISSUES IN IMPLEMENTING CLASSROOM-BASED ASSESSMENTS FOR LITERACY 31
Elfrieda H. Hiebert and Kathryn Davinroy

TEACHERS' IDEAS AND PRACTICES ABOUT ASSESSMENT AND INSTRUCTION 57
Hilda Borko, Maurene Flory, and Kate Cumbo

PARENT OPINIONS ABOUT STANDARDIZED TESTS, TEACHER'S INFORMATION AND PERFORMANCE ASSESSMENTS 73
Lorrie A. Shepard and Carribeth L. Bliem

Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics[1]

Roberta J. Flexer and Eileen A. Gerstner[2]

CRESST/University of Colorado at Boulder

INTRODUCTION

Let me set the stage for this paper by asking you to imagine that you are about to teach a basic statistics course. A team of cultural anthropologists and social constructivists offer to come help you decide how to assess the performance of the students in your class. Furthermore, they tell you they have novel ideas about what you should be teaching and how you should be teaching it. Your dean thinks it's a great idea and convinces you to "volunteer" to work with this team, to "welcome" the group, and to spend half a day each week for a year "discussing" how you should be teaching and assessing your course. The feeling you are now getting in the pit of your stomach might give you some indication of how our volunteer teachers might have felt about getting involved in this project.

---

1 This paper was presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April,1993.

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

2 We thank Abraham S. Flexer for his support throughout the project and for his editing of this manuscript. We also thank the team of graduate students for their many hours of work on the project, particularly the hours of sitting through meetings with teachers, transcribing tapes, and checking transcripts. They are Carribeth Blum, Kate Cumbo, Kathy Davinroy, Maurene Flory, Bernice Harris, and Vicky Mayfield. We give special thanks also to Pam Geist, a visiting researcher, for her very valuable contributions to the teachers and to the research team.

We are particularly grateful to the District administrators and personnel and of course to the hard-working teachers of Pine, Walnut, and Spruce Schools.

The project was prompted by a prior study that showed the negative effects on reading and mathematics understanding of high-stakes standardized testing at the elementary level (Flexer, 1991; Hiebert, 1991; Koretz, Linn, Dunbar, & Shepard, 1991; and Shepard & Cutts-Dougherty, 1991). It appeared that teachers focused their mathematics instruction on basic facts and computation; they also engaged in some questionable instructional practice under the pressure of preparing students to take standardized tests. As a result, students' performance on the standardized tests did not generalize to other tests of the same material. Other studies also described wasted class time and energy spent in preparation for the end-of-year standardized tests (e.g., Smith, 1989). At the same time we hear from many teachers that the nature of their instruction in mathematics is dictated by standardized tests and that as long as low-level skills and definitions are the primary focus of testing, that is what they *must* teach. By contrast, there are many claims for the benefits for instruction of performance assessment (Wiggins, 1989).

This project sought not only to free teachers from standardized testing but also to assist them in developing a system of performance assessment. The idea for the mathematics part of the project was to make teachers *free* to instruct as they were being encouraged by the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989; referred to in the rest of this paper as NCTM *Standards*) and to give them *support* to work on that instruction and on the assessment of their students being taught this way. The hope was that the refocus of assessment on performance, understanding, and higher-order thinking would, in turn, have a positive effect on instruction. We have evidence from Gipps (1992) that work with performance assessment (the UK's Standardized Achievement Tasks, SATs) had positive effects on instruction for a significant number of teachers. We were not so naive as to believe this would be an easy task. We know that teachers' initial attempts with performance assessment are fraught with problems. Badger (1992) reports the concerns of teachers involved with the Massachusetts performance assessment:

> The most frequently voiced concern about this type of evaluation has to do with time. 'If I am constantly evaluating, when do I find time for teaching?' Another concern refers to the amount of information that is collected. 'How do I record all this material and how do I report it? Do I have to make a case study of each student?'

The literature on teacher change in general (e.g., Richardson, 1990) and teacher change specifically in mathematics (e.g., Nelson, 1993) discusses the difficulties of effecting change, the sets of conditions conducive to change, and some success stories.

General issues of teacher change often concern: (a) the organization surrounding the teacher; that is, the culture in which the teacher finds herself or himself and the amount of support the organization provides; (b) the beliefs and knowledge of the teacher (in this case, about how children learn, about mathematics, and about the teaching and learning of mathematics); and (c) the active involvement of teachers and the acknowledgement of their control in making changes (Nelson, 1993; Richardson, 1990). These three categories are not disjoint and clearly interact; for example, the issues around a change that an organization might make that is not consistent with teacher beliefs belong in both category (a) and category (b), but the categories provide a useful perspective.

METHOD

This project sought to work on the content areas of reading and mathematics regularly (weekly) for a year with all the teachers at one grade level in several schools in one school district. We envisioned working with teachers to help them develop their own performance assessments and to support the changes the teachers wished to make in both instruction and assessment. We were interested in finding a school district willing to participate that had (a) standardized testing in place, (b) a large range in student achievement, and (c) considerable ethnic diversity. Such a district was found near an urban center and was selected for this study. The district population ranges from lower to middle socio-economic status. The research team decided to work with teachers of Grade 3, and only in schools in which a letter of application was signed by the principal, by all teachers at that level, and by the school's parent accountability committee. Three schools, with 14 third-grade teachers, were selected and matched with three comparison schools where data would also be collected. While all teachers were technically volunteers, it is possible that some were less enthusiastic than others to engage in the project. Some of the original teachers who volunteered changed grade levels or schools and were replaced by teachers who found themselves

involved in a project for which they had not volunteered; others may have been "strongly encouraged" to volunteer. By chance, all school personnel involved in this study are women (teachers, principals, district administrators), as are the researchers (faculty, graduate students, and visiting researcher).

The purpose of this paper is twofold: (a) to discuss some of the dilemmas and issues that arose in our first two terms of work with the participating teachers, and (b) to give a progress report on how teachers have changed their instruction and assessment as a result of the project. The primary data for this paper are the interactions about mathematics instruction and assessment that the research team had with teachers at each school. Project staff met with teachers from different schools each week to work on mathematics and reading. Schools alternated each week between the two areas. These meetings were audio-taped and transcribed. The data also include classroom assessments discussed at those meetings (student work on those assessments was sampled and copied) and transcripts of some parent-teacher conferences. The research team also visited each classroom once or twice, to observe, to do demonstrations, or to model individual and small group assessment.

The transcripts of the weekly meetings were analyzed, and a coding scheme was developed iteratively as it was applied to successive transcripts.

RESULTS

This section presents data about the mathematics part of the project and our inferences in three areas: (a) the dilemmas and issues with which teachers in all three schools struggled; (b) dilemmas or positions unique to a particular school; and (c) changes observed in the teachers in each of the schools.

Dilemmas Encountered in More Than One School

Many issues, concerns, and problems that arose in all three schools came to be called "dilemmas" by the project staff. We address those as a group and give some indications of how different schools handled them.

The issue of organizational support was clearly crucial to this project, for inadequate organizational support could effectively block any efforts toward change by the participating teachers, whether in assessment or in instruction.

In this project, there was very strong organizational support, and it was present at both the building and the district levels. Teachers and the research team had the support of administrators, curriculum specialists, and principals. That support came in both tangible and intangible forms: Teachers were aware that they had this support, and enjoyed many benefits, for example, extra planning time and budgets for materials they might need. The issue of organizational context was mentioned by teachers as a problem for only one school and will be discussed later.

The most difficult problem for teachers and the first to emerge was the lack of time, a universal problem of teachers dealing with change, whether about assessment, instruction, or other kinds of reform (Badger, 1992). Whether this issue is classified as organizational or teacher-related matters not. It is a serious problem. First, all of these teachers had their regular, ongoing classroom responsibilities to deal with even as they were trying to make changes in assessment and instruction, and several had professional or community projects occurring simultaneously. Second, we asked that each teacher work in both reading and mathematics, although almost every teacher asked if she could focus on just one of those areas at a time. The problem of time was presented in several different forms. Teachers did not have enough time to select and prepare performance assessments outside the classroom; there was not enough time in the curriculum to add new things, and their time seemed fragmented; the performance assessments simply took longer than the traditional assessments, and the teachers did not like to lose instructional time. Teachers in every school found that they were spending more time on the topics we were working on as a project and began to worry about all the other things they were responsible for teaching. They considered instructional and assessment materials from the project to be an *addition* to what they were teaching. They wanted to do what they had always done, and *then* add the new materials if there was time. These comments came from teachers at Spruce, but are similar to comments repeated in each school.

> T1: You know what I'm finding the most difficult in this whole assessment piece that we're going through, is having two subject areas to work on. I think I could be doing a better job if I was focusing on math or reading. . . . You know, it's just been a real management nightmare.

T2: Well, I think we're going to be a little bit slower because we're teaching this in a different frame of mind.

T3: I'm kind of scared that—I mean I don't think there's anything wrong going on, I'm just scared—are we going to get everything in because they have to learn money and they have to learn telling time and we are doing addition and subtraction and fractions and you brought up multiplication and sometimes we try to throw in division to get all this done when we are doing it so thoroughly. With this, there isn't anything wrong with it, it's just the time element, you know.

We and the district administration helped to address the time required outside the classroom in a number of ways: (a) we arranged to give teachers course credit for their work with us (while not freeing time, this gave teachers an extra benefit for spending their time); (b) the district provided teachers with a half day a month of released time to work on this project; (c) we reduced the amount of work expected of the teachers each week; (d) we staggered assignments for reading and mathematics, so that the teachers could focus on just one in a given week; (e) we supplied each teacher with a lot of resource materials so the teachers did not have to gather these themselves.

The problem of time within the classroom, we felt, arose because teachers viewed the new instruction and assessment as add-ons to their already full curricular schedule. We tried to help them see that instruction should take no longer if they did not try to do double teaching—the old and the new. We heard less about this problem as time went on and as teachers replaced some of their prior instruction and assessment with the new material or, in the case of multiplication, chose to use an innovative unit as their curriculum.

Teachers were learning what performance assessment was all about, and as good students, they expected to be challenged as they learned new things. There were five areas of assessment with which every teacher struggled:

1. what to assess
2. how to assess it
3. how to score the assessment
4. how to keep track of the results
5. how to report the results (grades)

In addition, most of the teachers also worked on instruction because the performance assessments were assessing a new curriculum that they had not been teaching (see below).

At the time of this study, the district was revising its mathematics curriculum and developing a framework consistent with the NCTM *Standards*. The district's "learning outcomes" for Grade 3 included communication, connections, number sense, geometry, problem solving, probability and statistics, measurement, patterns, algebraic methods, and confidence with mathematics. In order to decide *what* to assess, each team of teachers had to select a subset of learning outcomes from the district's curriculum guide on which they planned to focus. We suggested that they select a small part of the curriculum for our initial work. Since third graders typically learn about place value, addition, and subtraction in the first three or four months of school, each school selected those areas and some combination of the processes of modeling, explaining, performing mentally, performing with paper and pencil, estimating, making connections, and problem solving. Their first task was to build an assessment framework for the selected outcomes. The framework was constructed with levels of proficiency that distinguish students who had limited understanding about the topic from those who knew more, and those students from students who knew even more. The problems the teachers faced in developing an assessment framework were different for each school and will be discussed below in the section on individual schools' dilemmas.

The next task for each school was to decide *how* to assess the learning outcomes they had selected. It had been our initial expectation that the teachers would develop assessment tasks using as models the many examples the research team supplied. Indeed, at the start, most teachers were not ready to develop their own tasks, but they were willing to select tasks to use from those models. (One teacher from Spruce Elementary School designed her own tasks from the beginning.) Our expectation then shifted to having the teachers try out a variety of tasks to see which ones gave good information and matched the new kinds of instruction. *Our* perspective was still on having the teachers develop a custom-tailored assessment program that they would find useful, and, in the process, having them learn techniques that they could apply to developing other assessment programs. The *teachers'* perspective was on

assessing *their* particular students *this* year; they did not want to be in the business of designing assessments. Many teachers in the study have moved beyond a willingness only to select among choices to developing their own assessments. However, their purpose, reasonably enough, continues to focus on the information they derive about their particular students and not on which tasks make the best assessment tools by providing that information.

An interesting dilemma that emerged in at least two schools was that of giving pre-tests. Most of the teachers liked the chance to show the children and their parents how much the children had learned, and this could be done by showing huge gains from pre- to post-tests. There seemed no interest in using pre-tests to identify areas in which a child was already proficient and need not repeat material already mastered. The dilemma for some teachers was whether it is fair to a child to assess, using a pre-test, in an area that had not been taught, since many children found taking pre-tests an uncomfortable experience. As part of this discussion we probed whether or not there are some kinds of pre-tests on which a novice could still perform. We have had several discussions about this dilemma, and it continues to be a topic of conversation.

The issues on the topic of scoring assessment tasks were not controversial but still caused some struggle. Teachers did a lot of work on designing scoring rubrics for explanations in problem solving, using student work to help develop the rubrics. They wanted a general form that could be applied to many tasks, not task-specific rubrics. They used their project-related work on developing criteria for a good summary as a basis for establishing their own criteria for a good solution and explanation to a problem. Also, as they had done in the reading side of the project, some teachers asked the children for their ideas about what makes a good explanation, and as a class process, developed a set of criteria. These were then posted in the classroom for students' reference as they worked on problems. One teacher staples a reduced copy of the criteria to each problem a student has solved so that students can understand the basis for their score. Many different rubrics were designed, some of which incorporated both the answer and the explanation into one score and some which gave two separate scores. Teachers are currently applying the rubrics and refining them. (See Figure 1 for some examples.) It is not yet clear if the more elaborate rubrics will prove too

A.
4 = Good explanation. Right answer
3 = Good explanation. Wrong answer.
2 = Right answer. Confusing/incomplete explanation.
1 = Wrong answer. No attempt at an explanation.

B.
4 = Correct answer. Clear, correct explanation written out. Correct spelling.
3 = Correct answer. Clear process. Clear but not as involved explanation.
2 = Correct answer. Poor explanation.
1 = Wrong answer. No explanation.

C.
4 = Correct answer. Tells a lot. Well written.
3 = Correct answer. Doesn't tell as much.
2 = Close answer. No sense.
1 = No answer. No explanation.

D.
Answer
4 = Correct.
3 = Almost correct.
2 = Reasonable.
1 = No sense. No correlation to the problem
0 = No answer.

Explanation
4 = Logical. Tells how they got the answer. Thorough.
3 = Solid. Essentially correct. Maybe left out a component.
2 = Vague, unclear. Picture doesn't match (if there's a picture).
1 = No sense.
0 = No explanation.

Figure 1. Examples of scoring guides for problem-solving tasks. A, B, and C use a single score for both answer and explanation. D uses two scores.

unwieldy for regular classroom use; at this point, teachers with simpler schemes seem to be using them more often.

As the teachers in Badger's study (1992) noted, there is a lot of record keeping that goes with performance assessment. All the teachers found

record keeping to be difficult because we were asking them to gather evidence about what children can do from sources different from worksheets and end-of-chapter tests. It meant gathering evidence on children's thinking from many sources, for instance, observations as children worked with manipulatives, verbal answers to specific probes from the teacher, responses as they played instructional games, contributions in discussions, and written explanations to tasks of problem solving and computing. Some of the record keeping meant having a place to put the scores from the rubrics they had developed. It also meant having a way to keep track of how well a student could model a computation, like adding two-digit numbers with regrouping, or how something he said in a discussion rang a warning bell about his understanding. So it meant having a place for short anecdotal notes.

It is in the job description of elementary teachers that they watch what their children are doing, and most of the project teachers felt they had a pretty good idea of each student's level of performance. We were asking them to keep track of their understandings in a more formal, that is, systematic and written, way. Some teachers felt this was needless because they had it all in their heads, but all were willing to try it. Some responses from teachers are:

> T1: I think it's been, it's good information and it's really helpful, it's just kind of overwhelming with both [reading and mathematics].

> T2: It's real time consuming and I also found as I was doing it I found that we needed some sort of a code to match it, so that if they're working on it, or if you really think they've got it or if they're having a lot of difficulty you can mark that somehow.

Many felt the record-keeping process was unmanageable at first, but all have persevered with designing and trying out ways to do it. The teachers who use whole group instruction found the task particularly onerous and management became an issue. What were they to do with the students who were not being observed and assessed? Those who typically used small group instruction had an easier time with the management issues, and their primary task was to find a workable way to keep track of the children.

The 14 teachers in the project seemed to generate dozens of schemes for keeping track, many of which were discarded after the initial trials. Their common characteristic was boxes, small ones for checks or larger ones for

anecdotes, and every students' name on one page. Only at the beginning of the project did some teachers try to use a scheme that had one sheet per child with a recording to show what the child had done in each area. That method proved to be too awkward for daily use; instead, there was a move to one sheet for the whole class. In some cases, one assessment activity was recorded at the top, in others two, three, or four, with notes next to each child's name for each activity. Some teachers adapted for mathematics record keeping a form the district had used for Writers' Workshop. It had boxes across the top with days of the week to record the primary activity for the day. Below that was a matrix of boxes, each with a child's name and each about 2 square inches, for notes about the child's performance. All teachers continue to work on record-keeping plans that will be informative but efficient and reasonable enough to use daily in classrooms where lots of things are going on. Generally, keeping notes about individual students proved a chore for all the teachers, and most of those who persevered with it used abbreviated systems of checks, pluses, and minuses. Teachers who wrote more extensive notes tended to do that as a joint project with a member of the research team.

What can be said about report cards is that all the teachers in the project are dissatisfied with them. (Is that a universal truth for all teachers everywhere?) The district is currently working on developing new forms, and the dilemma is whether or not to use a form in which a child's performance is compared to a standard. Each school has its own form, and while different symbols and descriptors are used, grades are given essentially on a three-point scale of superior, satisfactory, and needs improvement. Timed test scores are also reported directly, which, as the teachers point out, imbue those assessments with considerable significance. The teachers have asked for assistance in redesigning their report cards, and that is in the queue for a future discussion.

Dilemmas concerning changes to instruction arose when teachers first looked at models of performance assessment and said that they were not suitable for their students who had not been instructed with that kind of assessment as the goal. Since the examples of performance assessment had been selected to fit the district curriculum, some of the teachers decided to make some changes in their instruction. Others continued to question parts of the district curriculum (see sections on individual schools).

All teachers, even those at odds with aligning assessment to the district curriculum, requested and seemed pleased with ideas for instruction that would be more in keeping with the NCTM *Standards*. There were fewer requests from Walnut Elementary School, but the same materials were prepared and delivered to all three schools. Every teacher used at least some of the ideas that the research team provided. Some of the teachers were already using many of the same kinds of activity. We tried to point out that the same activities could be used both as instruction and as assessment, as Badger (1992) so aptly states:

> . . . in more activity-oriented classrooms, no differentiation exists between the types of tasks that are used for instruction and those used for evaluation. Both should be interesting, challenging activities that reflect important themes in the mathematics curriculum. The difference occurs not in the kind of task but in the role of the teacher.

In many of the conversations we had with teachers we discussed the differences between assessment and instruction and how interrelated and intertwined they are. Some teachers looked at the tasks presented as "fun" things to do in the classroom but saw little value in them as sources of information about how children understand a piece of mathematics.

Individual Schools

This section characterizes each school in turn and presents some dilemmas and issues unique to each school or a position the teachers of that school took on a common problem.

Pine School

Characterization of Pine School. Pine School prides itself on being at the forefront of innovation in curriculum and instruction. Whatever is happening in elementary education, Pine is doing it. The third-grade teachers are aware of their position in trying new curricula and struggling with new ideas, and it is no surprise that the school volunteered for the CRESST project. The teachers talk about how exhausting it is to work on many new programs, but all seem eager to be involved in each new project, and they engage in them with energy and professionalism. Most of the teachers are well seasoned and have taught at the third-grade level for many years. They are a cooperative

team and support each other strongly; they particularly support the least experienced teacher. They work well together in spite of divergent ideas about teaching mathematics. They eat lunch together daily and often discuss their classes and plan together during their short lunch break.

Manipulatives are available in the school, for example, base ten blocks and pattern blocks, and the teachers use them to varying extents. Two teachers have taken the district's version of Marilyn Burns' *Math Solutions* course and use manipulatives regularly. One of these teachers has a goal for the year to work on problem solving and written explanations with her students, and she began to do that at the start of the academic year. She has a chart of problem solving strategies in her classroom, and encourages children to select an appropriate strategy to solve each problem she assigns; she then asks children to share the strategies they used to solve each problem. The second of these teachers also stresses problem solving and gives instruction on strategies. By the end of the fall term, the other teachers had incorporated problem solving into their curricula as well.

All the teachers use a popular commercial textbook, along with supplementary material from resource books in the school's collection. Most, though not all, instruction is addressed to the whole class. One of the teachers mentioned above divides her class into two groups so that she can teach one group while the other does practice exercises from the text. Instruction is didactic, clear, and carefully planned. Occasionally, teachers use learning centers which they plan, prepare, and use together, moving all third graders through each center. Pine teachers' team effort in planning and sharing materials is one solution to the time dilemma, a solution already in place in this school.

Their major assessment at the start of the project was multiple-choice chapter tests. Some of the teachers gave the text's pretest at the start of a chapter and the posttest at the end, so that children and parents could see the progress that had been made. The school requires that every teacher give timed tests on addition and subtraction facts and, later on, multiplication facts, so these were also part of assessment. Most of the teachers send drill-and-practice sheets home for students to work on with parents, and at least one teacher also sends sets of timed tests home for parents to use, score, and record. One reason they do this is to have more class time available for

instruction of concepts and problem solving, rather than using it for drill and practice.

Dilemmas at Pine School. In the first six months of the project, issues and dilemmas arose for this group of teachers in almost every category. Their manner of dealing with dilemmas was to engage in lively debate with the research team and each other and to argue their positions forcefully. Rarely in these discussions did all the teachers take the same side. The team leader often stated a position that might be unpopular with the research team but did so to represent the views of several teachers. The more experienced teachers were the more vocal in these discussions.

The group used the district curriculum guide as they worked on an assessment framework for the performance assessments of the first semester. Because we had advised them to start on a small section of the third-grade curriculum, they selected place value, addition, and subtraction. The first issue that arose was the place of problem solving, which they saw as a separate topic that comes after children learn basic skills. The research team convinced them that problem solving belongs at every level, along with other processes like modeling and explaining.

While the teachers nominally accepted the district's framework for their curriculum, they interpreted it differently from the research team, and questioned some of the outcomes, for example, the first outcome of "communicating mathematical ideas."

> T1: I still question why kids have to write this all the time. I still question that. I still would like some real rationale for that.

This teacher felt that children who are less verbal and might be brief in their responses might be graded down; she said she didn't see why a verbal explanation wouldn't suffice. A second teacher objected to asking for explanations at all because they might penalize the child who did not have to figure a problem out, who just "got it" and cannot say how. The child's ability to get the correct answer is the important assessment, she asserted.

Under the heading of number sense in the curriculum guide is the outcome that a student "uses mathematical concepts and arithmetic operations with understanding." The teachers and researchers had many discussions about teaching for understanding.

T1: But do you think, _____, that all students will totally understand?

R: Well, that's what you're pushing for. That's what you're trying to get at.

T1: But don't think you think that some children will only learn it by rote?

Later in that same conversation, after describing a child who must count each time to add, the same teacher says:

T1: . . . a child like that maybe we're better off just teaching him how to add and subtract on paper the traditional way, because that child may never until he's maybe 30 understand what he's doing. See, I'm not sure that understanding has to come before doing it. I think many times doing it on pencil and paper, later then will help you understand it. See, I'm not so sure that understanding has to come first. Because I think some children aren't capable of understanding.

But this same teacher then agrees with another who mentions using strategies to add, and says she teaches things like, "if it is adding nines then add ten and subtract one." She goes on to say, "there are very few facts that they really have to just memorize." She also takes the position that one need not mindlessly follow an algorithm, that starting with the tens column in an addition problem is a good idea if you are estimating. Most of her colleagues disagree totally—in adding you *must* start on the right. "Estimating is [a] different, see that's not adding."

As the teachers discuss this point, two recall that children in their classes are able to add two large numbers by first combining tens, for example, 28 + 47, "So they would say 20 plus 40 is 60. And then I know that 8 and 7 is 15. So that makes 75. I was just amazed." As the teachers discuss and reflect, they have a chance to revise their opinions to become more flexible in their thinking about algorithms. Some, though not all of them, do. But the process isn't one of steady progress. Just after this part of the conversation came:

T: But I guess my problem with that is that some kids get confused so easily, that by the time you have shown them five different ways of doing it, they would be totally confused. I mean I think there's a number of children, maybe three in a class, that just don't know.

Another issue arose in October during the selection of instruction and assessment tasks to go with the framework the teachers were using. The team had worked diligently and efficiently with materials they had gathered from

the research team, their school collection, and the district's resource center. They had made excellent choices of instructional activities, but misinterpreted what a member of the research team had said about assessment activities. They thought the assessment tasks had to be quite different from the instructional tasks, and so they selected very traditional workbook pages for their assessments. They were satisfied to use these even after they were told of the misinterpretation, because they felt that after children had used manipulatives to gain understanding of a process, then the "bottom line" question was, could they do the process with paper and pencil. After all, they argued, they won't have manipulatives available when they're adults. Their inclination was to go to multiple-choice and paper-and-pencil assessment each time, claiming that the ability to perform the operations showed a child had passed through the levels of understanding to a more meaningful result. They also felt more comfortable that a child had done his or her own work when it was privately produced on paper, as opposed to displayed on a table with manipulatives that all could see. The issue was not really resolved by discussion but by the researchers requesting that they use tasks similar to the instructional tasks for assessment along with their traditional forms of assessment.

Another issue for Pine was the need to prepare children for each assessment task, that is, to teach them every part of a problem they were asked to do. The teachers seemed to feel that children cannot figure things out if they haven't been shown how to do them first, that they can't hypothesize, conjecture, and test their hypotheses. All the teachers had these concerns in the fall, and they surfaced in discussions about problem solving, when they were encouraged to present some tasks that the children would never have experienced. One teacher said, "All of these are separate things that you would have to teach them." Another suggested that you would have to teach the strategy at some other time in the day. Only recently are they seeing that children can make discoveries for themselves and can investigate problems of their own devising. For one teacher this happened during their unit on multiplication, for which they were using Marilyn Burns' *Math by All Means*, when two children decided to make up a rule for finding prime numbers (numbers that could be represented by only two rectangles, 1 X the number and the number X 1). How interesting that these children had

stumbled on a problem that has concerned mathematicians for hundreds of years!

Walnut School

Characterization of Walnut School. Walnut is a fairly new school that is well equipped with manipulatives and resource materials. Its teachers pride themselves on activity-oriented instruction. Much of the classroom work in third grade is done with children seated in groups. During the school year the third-grade students work on several large, integrated projects, and mathematics is taught in the context of these units. The teachers also incorporate mathematics from the daily events of the classroom, like lunch counts and the calendar. Several of the teachers have been teaching at the third grade for seven or more years. Most of the experience of another teacher is at lower grades, and another teacher was new to third grade this year. Three of the teachers each have more than 17 years' experience. They are an energetic and focused group that sometimes works as a team, but more often in two subgroups. Several of the teachers have been to workshops on teaching mathematics and have large personal libraries of resource materials that they were happy to share with the teacher who had never taught third grade before.

The philosophy of the teachers at Walnut is very close to the newly revised district curriculum, although their assessments in the past have been fairly traditional, for example, worksheets, or a set of four computational examples and one non-routine computational problem each week.

Dilemmas at Walnut School. Walnut teachers' first dilemma arose when they tried to establish an assessment framework. Their initial goal was to work on "reasonableness" with respect to computation, that is, the ability of students to look at an answer to a computational problem and decide if that answer makes sense. They were quite willing to broaden the application of reasonableness to other mathematical areas and began to build an assessment framework that had elements similar to those at the other schools: the content of place value, addition, and subtraction, and the processes of modeling, translation, explaining, and problem solving. One teacher said, "Things are so intertwined that you can do problem solving with place value." Initially Walnut was the only school that viewed it that way. Teachers tried to put the content on one continuum and "reasonableness" on another, with the

processes crossed with both. Their final framework did not have "reasonableness" either as a dimension or a category, but it was their intent to use assessment tasks that probed students' abilities in that area. We had several conversations about the difference between a continuum and a rubric, and they decided to have a rubric for reasonableness.

They worked diligently and probably too intensively on the framework, in that they divided the content dimension into smaller and smaller categories, as if they were defining each task, rather than dealing with a broad area of accomplishment. When they finished with the framework, they felt it was uselessly fragmented and did not fit their teaching, particularly when they were involved in the large integrated projects. The research team agreed, particularly the mathematics educator who was getting more nervous all the time about all the little boxes they were producing. It is ironic that teachers at this school were already incorporating processes like modeling, explaining, and problem solving into their curriculum, but when they wrote it out on paper, the framework looked too onerous.

The framework may have influenced them to spend much more time on place value than they might otherwise have done because it had been partitioned into very small bits. But they also felt that spending a long time on place value was going to give their students a sound foundation for addition, subtraction, and multiplication. Of course that decision in September and October caused them panic in April when they realized how much of the third-grade curriculum they had not "covered."

Walnut's dilemma on how to assess had to do with the integrated instruction that they valued. Since they did not know how to assess such instruction, they assessed computation and computational problem solving in a weekly assessment, called Big-5. Students were given a computational exercise to do each day and possibly one involving some problem solving on Friday. They also gave the school-mandated timed tests, but put no emphasis on preparing children to take them. At least two of the teachers reported the results on timed tests during parent conferences. They did not ask children to write very much as part of their mathematics assessments at the start of the year, but began to do more in the winter semester.

Teachers at Walnut also struggled with the pre/posttest problem; they liked the idea of showing growth, but did not want to present their children with a painful pretest.

The issue of how to assess the mathematics in the integrated units remained a problem. We asked what observations they had made of children during their mini-society, a unit in which there were many rich mathematical opportunities. As part of their mini-society, students ran businesses and produced or bought items at wholesale prices that they later sold in a school store. Students had to take loans at 10% interest to buy the items, had to price them, sell them, and calculate their profit. The teachers were frustrated at missing opportunities for assessment, beyond some observation of practice at making change.

Walnut teachers have continued to ask for checklists for observation and for help with how to do it, even for daily work outside of the integrated projects. The research team was fortunate to have a visiting researcher join it who could work each week with the teachers at Walnut. She has been modeling assessments of individuals and groups of children and holding conversations with the teachers regarding what she is learning about their children.

Spruce School

Characterization of Spruce School. The teachers at Spruce Elementary School have a wide range of teaching experiences—from a few years to about 25. As in the other schools, one of the third-grade teachers is new to that grade this year. The school seems to be fairly traditional. The teachers there value basic skills along with thinking skills, and they are eager to give their students a sound and thorough grounding in the basics. Spruce teachers use primarily whole-class instruction with a popular textbook and additional worksheets; they have a few manipulative materials available, though only limited supplies of each. These teachers' confidence with mathematics and its teaching was not as high as it was in other subject areas. Their assessments in the past have been end-of-chapter tests, worksheets, and timed tests.

As a group, the teachers were quiet in meetings and let the researchers do most of the talking. Perhaps they were overwhelmed by the whole project. Because they rarely objected publicly to the ideas of the project, and chose not to argue about points with which they may have disagreed, it was not always

clear how they felt and what their problems were. As at the other schools, several teachers expressed the desire to keep the number of curriculum goals that we would work on to a minimum. One said, "I would really like to keep this narrow until I kinda get the feel for what we're looking at."

Most of the time, these teachers tended to work as individuals, but they always seemed very willing to share instructional or assessment tasks. They had little time together outside of their weekly meetings with us to work on the project as a team, except for the half day each month that they were released from teaching. They all took a class during the first term on using hands-on instruction given at their school by the school's half-time mathematics specialist. This specialist was also available on a regular schedule to demonstrate and teach with them in their classrooms. The presence of the specialist in the school had some positive, but perhaps also some unexpected and problematic, consequences (discussed below). The teachers learned a great deal from her and received more resource materials to use. She also reinforced with them many of the project's ideas.

Dilemmas at Spruce School. It seems, in retrospect, that two things might have been happening as a result of working with the specialist and the project: (a) they were being pulled in a direction opposite to that of the other teachers in the school and of their students' parents; and (b) the teachers may have been losing control of their instruction. In December one teacher expressed this idea in a powerful way; the others concurred:

> T: It's real frustrating because I know what the thinking is and I know what pretty much what we're supposed to be doing. But then I was talking to a fifth-grade teacher the day before yesterday and she was saying how the kids don't know their facts and they can't do their computation skills. It's like we're being geared to do problem solving with the kids and all that and then teachers in upper grades are upset because they're coming into them and not having the computational skills that they think they should have. One teacher does math time tests and we hear, "No we shouldn't be doing math timed tests, that's not a valid way for kids to learn their facts." It's like being pulled in two different directions. And we can teach the problem solving and, at least we're trying to be able to do that. Not all people believe that that's the way- what we should be doing and then we send our kids up to them, and it's like, "Could this child do their timed tests when they were in third grade?" Do you know what I mean? Don't you guys feel like that, like you're being pulled in

two different directions and then parents come in and say "I don't understand why my child doesn't bring home 25 addition problems every night to work on, what good is this going to have them do to count the legs on this animal."

These teachers were finding it very difficult to reconcile other teachers' expectations, parental demands, school goals, and their own beliefs with the ideas being presented by the researchers and the mathematics specialist. While the teachers were supported by the principal and mathematics specialist to make changes within their school, it appears that the pressure they perceived from other teachers was very much in another direction. Recall that one of the issues in the literature on teacher change is the importance of the surrounding organization—its support and context. The Spruce teachers were having a very difficult time within the culture of their school when nominally it appeared that they had the district's and school's support, but in practice, they felt subjected to countervailing values.

An issue for teachers making major changes is that of control; teachers are in control of their change, and that must be acknowledged (Richardson, 1990). As these conflicting pressures continued, the Spruce teachers seemed to become less decisive about their curriculum, and less in control of their instruction. Perhaps this loss of control had some effect on their interest in making other changes. By the winter semester, they were having a difficult time planning what their curriculum would be for the remainder of the year. The mathematics specialist was scheduled to be in their classrooms several days a week for about a month, and she seemed to take over the unit on multiplication. As in the other schools, they were using the Marilyn Burns' *Mathematics by All Means* unit, but the specialist was introducing it, planning all of the instruction, and teaching the classes with the assistance of the teachers. When asked what they did in mathematics on days that the specialist was not in their rooms, they joked that they didn't do any mathematics.

The dilemma of what to assess (and even what to teach) was very much a problem for the Spruce teachers, although they did not complain or argue about it. One teacher seemed to speak for the group when she said:

T: I personally, I still feel like I need a balance of both. I don't want to do all problem solving everyday, this kind of problem solving. And I don't want them to

do all pages out of their books everyday. But I do think for them to survive I think they need a balance, and I want them to be able to do some thinking skills, but I also, if they go to fourth grade next year and the teacher says you need to do page 36, 1 through 25, I don't want them to look at each other and not have a clue on what they would do with something like that ... not know how to put a heading on their paper or write their numbers so that they can be read by other people. I think they need those things from that kind of practice no matter how well they know their facts from playing cards. I just think there needs to be both. I think they need to be able to write problems on paper and have somebody else be able to read them.

A major concern for the Spruce teachers was the role of the timed test, a topic that came up at many of our meetings. The teachers thought that, according to district policy, all teachers were required to give their students timed facts tests. Each school had a different version of who required these tests, and it was half way through the year before we found out that they were not required by the district, but that each school set its own policy with respect to the administration and standard (e.g., 80% of 100 facts correct in five minutes). The Spruce teachers seemed to take these tests more seriously than teachers at the other schools. They gave each other advice on how to improve scores and made many comments about the tests, including one that pointed to the inability of other kinds of assessment to prepare students for timed tests.

T1: Try this, try telling them they have to finish it.. And just write their time down. Keep track of their time. And they have to finish it. It's amazing how much faster they get.

R: Is the timed test a policy of the school?

T2: Yes. We've talked about it the last couple of years. It has been a topic of discussion but as of right now it's on the report card and our kids are expected to do 100 problems in 5 minutes. ... So you're thinking that it shouldn't really carry that much weight with the time limit on it?

T2: The fifth-grade teacher, how can I say to her, "We just don't feel that math timed tests are important for kids anymore?"

T2: And if they're supposed to get, I mean if the school's goal is to get 80%, 100 problems right in 5 minutes, how else can you do that except to give them a timed test? How else can you get that data?

T3: And I think the timed test is used to see how they're doing and this stuff is working so they can improve their score. Well, I use the timed test for assessments, is what I'm saying.

When discussing record-keeping for the project, the Spruce teachers were no different from the other participating teachers. All of them complained about the amount of time required to keep records. For example, one teacher, in talking about making individual observations of children working, said:

I think it's been, it's good information and it's really helpful, it's just kind of overwhelming with both of them [math and reading].

Another said:

It's real time consuming and I also found as I was doing it I found that we needed some sort of a code to match it, so that if they're working on it, or if you really think they've got it or if they're having a lot of difficulty you can mark that somehow.

They felt the record-keeping process was unmanageable at first, but several of the teachers invented good forms to be used while doing observations, and one of the teachers was able to use hers effectively. We shared the forms with teachers at other schools.

Although the project dealt with assessment, a great deal of time was spent discussing instruction with all the teachers. The Spruce teachers used traditional teaching methods, but wanted to learn about other ways. They focused on learning new methods and using new materials, not on how instruction is related to assessment. We spent a lot of time gathering materials and discussing how to use them in instruction or as assessment. Calculators, whose use was encouraged in the district curriculum guide as it is in the NCTM *Standards*, were not used because, as one teacher explained:

T: ... and that's why we don't give them calculators today until they learn it, learn the process and the understanding of the process because they can take a calculator and punch in the buttons and get the answer anyway.

There was some limited use of manipulatives, and as teachers learned what could be done with manipulatives, they used them more. They seemed to be interested in obtaining more manipulatives, but this did not occur despite the nominal availability of funds; possibly because of reliance on the mathematics specialist who had a supply of materials she used during her instruction.

The dilemmas at Spruce Elementary School seemed not to apply to all teachers equally. One not only used many of the supplied tasks, she began early on to invent her own for both instruction and assessment. For the most part they were exemplary performance assessments, and she was willing to share them with the other teachers at her own school as well as teachers at other schools. She also devised new continua as she moved into new topics in the curriculum, and these reflected the spirit of working on both content and the processes of modeling, communication and problem solving. She designed a Likert-type self-assessment, using happy faces for markers, that the children filled out for her.

Changes

What changes did we see in the teachers' instruction and assessment? There were both school effects and individual changes. The good news is that we saw lots of change (albeit some in small steps), and it was in directions that reflect the NCTM *Standards*. The discouraging but expected non-news is that change is slow and non-linear. And the more profound the change, the slower. Had we not gotten the latter results, you would have good reason to doubt our story.

Changes at the Level of the Schools

The Pine School teachers changed their assessment and instruction the most. They came from questioning all of the process strands beyond showing paper-and-pencil computation to a quite different view of assessment. In an April meeting, for example, they were asked what they planned to teach about geometry. The team leader immediately said, "Well, we first have to ask, what do we want kids to know?" Then she and the other teachers made suggestions well beyond the level of the identifications they had worried about in October. Their ideas for assessments also involved more active and engaging tasks, and they just expected that observations of the children working would be a good source of information. The Pine teachers also developed a unit on probability from a variety of sources—materials the research team brought, books in their resource library, and materials from the district resource collection. They spent one session with the research team discussing ideas about probability, its pedagogy, and its assessment. Pine's instructional activities were solid,

interesting, instructive, activity-based, and attractive to the children. Their assessment for this unit could be considered alternative, including assessment of problem solving and communication. As stated earlier, the instruction that followed the probability unit was Marilyn Burns' multiplication unit, and its instructional design called for lots of activities for small groups. It is impossible to say whether our assessment project had anything to do with their selection of this unit. One of the teachers on the team had seen the unit in December, told the others about it, and they had their books in January. They supplemented the suggested assessments in the Burns unit with a set of tasks they developed that was much more open and conceptual than the end-of-chapter tests they had been using, although it was still light on problem solving and communication.

The Pine teachers continue to use the end-of-chapter tests, but they now supplement those with other assessments that involve more conceptual understanding and higher-order thinking. They also like representing a concept in many different ways; for example, they decided to include arrays of dots in their next assessment of multiplication to see if children could extend the idea from the arrays of tiles they had been using.

The Pine teachers were willing to eliminate some topics from their usual curriculum (e.g., division into a three-digit number) in order to do more work on problem solving and to do geometry more thoroughly.

Pine teachers frequently expressed surprise at how well the students were doing, how much they knew, and how much they enjoyed math. The teachers see evidence that the children are getting better at problem solving and at writing explanations for their solutions to the problems. In the first months, these comments were often accompanied by doubts of some kind about the assessments; we don't hear those much any more. (Perhaps they got tired of telling us.)

The conferences the Pine teachers had with their students' parents included a discussion of the new assessments and how important these are and how much they say about what their child knows and can do. We take it as good signs that the teachers value the performance assessments sufficiently to discuss them with parents and that the assessments were presented positively and convincingly. It is hard to say at this point whether these

changes represent the beginnings of fundamental change in beliefs, or less substantial change in instruction. It is possible that some of the Pine teachers are at the beginning of an epistomological shift. We believe that belief and practice can be causally related in both directions, that is, that the shift in practice may lead to a shift in belief which can lead to further shifts in practice. As these teachers are reinforced by seeing that their students are learning more, for instance, seeing their students talk about the connection of multiplication to geometry, they may change their ideas about how children learn mathematics.

The Walnut teachers did not change their instruction as much as the Pine teachers did, but that was because the Walnut teachers were already using lots of hands-on, small group activities. For example, Walnut also used the Marilyn Burns multiplication unit, but they had used it last year as well. When asked in December what they were doing differently in math, one teacher replied:

> T1: In math, I think for me the biggest thing is to get the kids to verbalize their thinking. Not just to have a strategy and apply it but then, then to regurgitate the strategy and what they were thinking.
>
> R: To recount the strategy.
>
> T2: Well, yeah. I mean and to just, to go back. I think there have been times that we have said, "Tell me how you got that," and the kids explains that and we say, "Great." This year we say, "Tell me how you got that and put it here (on paper)," and I have proof to show parents and to show me.
>
> R: So tell them to write [it] down?
>
> T2: I know that I am doing more writing [by students]. I did some before but I have to say it was very limited.
>
> T1: Also something I'm really encouraging with my kids is to be flexible, that there isn't one way. Today we solved a problem and we got six different explanations of how you could have possibly solved it. In my mind, math has been, in the past, right or wrong, and I'm really trying to encourage them to think flexibly, to be flexible in their thinking that, well if it didn't work this way I could try this, or if it worked this way could it work another way? Could I look at it from a different avenue?

Walnut's changes in assessment involved learning more about how to observe and question children one-to-one or in small groups and how to take note of what individuals are doing in large groups. In the multiplication unit, they used the assessment tasks suggested by the text and, in addition, designed an assessment that was alternative in nature. Several of the tasks on the Walnut-designed assessment probed the concept of multiplication and the commutative property; several also required explanations, including a traditional textbook-type problem.

The changes in some of the Walnut teachers seemed more subtle and deeper. They seemed to be struggling further to align their beliefs about pedagogy with their practice. They were interested in and eager to discuss some very fine points, like whether a particular rectangle should be referred to as 2 X 3 or 3 X 2. They were willing to go beyond the idea of "which is correct" to how children could look at it in different ways and what the implications of that might be, and to consider whether one was conceptually more sound than the other. The change for some Walnut teachers was not so much in adopting different beliefs, but in moving further along the lines of the beliefs they already held.

Teachers at Spruce incorporated into their classrooms many of the tasks supplied by the research team. They requested instructional ideas, were willing to use them, and were pleased with the results. Spruce teachers were aware that their students seemed to understand mathematics better and were enjoying it more. One of the teachers said,

> T : They sure love math I tell you... You know they love these games we're doing. They just understand it so much better and I'm doing a better job of teaching it.

The same teacher viewed their role as one of piloting the program. She felt they had learned a lot and they would do much better next year.

One of the teachers at Spruce seems to have made major changes in both her instruction and assessment. She seemed intrigued by the ideas of the project, was willing to try some new things, and produced her own performance assessment tasks, many quite fine. These tasks often incorporated computation, whether they involved problem solving or spatial relations. She also produced assessment frameworks for each new topic and developed a Likert scale for students' self-assessment. It is too early to say if a

change in epistomolgy is occurring, but it is clear that she owns the ideas of the project and uses them to extend her own teaching.

The culture of this school may have made it difficult for teachers at one grade to make significant changes. Even if they all agreed that they wanted to change, they feel pressured by teachers at higher grades and perceive that their parents don't want them to.

CONCLUSION

The aim of this project was to help teachers adopt performance assessment in their third-grade reading and mathematics classes and to free them from the constraints of standardized testing. As expected, the preliminary results were mixed, but hopeful.

This paper looked at the problems, issues, and dilemmas teachers encountered as they attempted performance assessment in mathematics for the first time. These issues were primarily in the area of beliefs and practical teaching knowledge. For some of the teachers the real struggle was not in learning how to use performance assessment, but in believing that is was a useful thing to do. The dilemmas most disturbing to them were those that focused on what is important to teach (and therefore assess) and how children learn. For those teachers whose beliefs about those issues were at odds with those of the district's new curriculum and performance assessment, the issues were more profound and the teachers' initial changes more superficial. Even so, some teachers appeared less disturbed by dilemmas that struck at core values. The changes those teachers made, even if not profound, may be of the kind that will eventually cause their beliefs to change: As they see positive effects in their students from a few changes, they may be reinforced for the changes and continue to do more and more until they eventually shift what they believe. The teachers whose beliefs were consonant with the new views appeared more comfortable with their dilemmas and were able to focus on more fundamental changes.

Was the greatest change seen in schools initially closest in philosophy to the district's curriculum? . . . farthest from the district's philosophy? Although Walnut's philosophy is most closely aligned with the district's curriculum, in the first term this school produced less in the way of

performance assessments than either of the other schools. Perhaps they were doing more that did not get assessed, for example, in their integrated unit; or they might have seen the tasks suggested by the research team as an intrusion on an already rich curriculum. In that first term, they tended to assess mainly computation, adding a non-routine, computational problem-solving task each week. They did little writing in mathematics assessment until the second term. In contrast, Spruce's mathematics folders were packed with assessments they had tried, even though they appeared to be the least enthusiastic about the district's new curriculum and the new assessments presented by the research team.

If change is measured by how well the ideas about performance assessment become integrated into the teachers' repertoires, then we cannot look at schools, but must think about individual teachers. Each school had a teacher or two or three who appeared to be engaged by performance assessment. The teachers in this group fell into two categories: (a) teachers grappling with ideas at deeper levels of belief and using them in their general movement toward teaching from a more constructivist perspective of learning; and (b) teachers who adopted the ideas into their current belief system, revising them to make them their own. Some of the teachers who changed the least may have felt they currently had quite adequate or even superior teaching programs and saw no reason to change. Others found themselves in too much disagreement with these notions; the ideas did not match their current views and therefore proved of little use or value. Others continued to be overwhelmed by trying out new instructional and assessment ideas in two major areas of the curriculum.

Given that teachers were in very different places in their beliefs about mathematics and about teaching and assessing mathematics before this project began, it was no surprise that there were great individual differences in their dilemmas with performance assessment, their reactions to those dilemmas, and the changes they made as a result of the project and their struggles with the dilemmas.

REFERENCES

Badger, E. (1992). More than testing. *Arithmetic Teacher*, 7-11.

Burns, M. (1991). *Math by all means.* New Rochelle, NY: Math Solutions and Cuisenaire.

Flexer, R.J. (1991, April). *Comparisons of student mathematics performance on standardized and alternative measures in high-stakes contexts.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Gipps, C. (Ed.). (1992). *Developing assessment for the national curriculum.* London: Kogan Page.

Hiebert, E.H. (1991, April). *Comparisons of student reading performance on standardized and alternative measures in high-stakes contexts.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Koretz, D.M., Linn, R.L., Dunbar, S.B., & Shepard, L.A. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Nelson, B.S. (1993, April). *Implications of current research on teacher change in mathematics for the professional development of mathematics teachers.* Paper presented at the annual meeting of the National Council of Teachers of Mathematics, Seattle.

Richardson, V. (1990), Significant and worthwhile change in teaching practice. *Educational Researcher, 19*, 10-18.

Shepard, L.A., & Cutts-Dougherty, K. (1991, April). *Effects of high stakes testing on instruction.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Smith, M.L. (1989). *The role of external testing in elementary schools.* Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 71*, 703-713.

DILEMMAS AND ISSUES IN IMPLEMENTING CLASSROOM-BASED ASSESSMENTS FOR LITERACY[3]

Elfrieda H. Hiebert and Kathryn Davinroy

CRESST/University of Colorado at Boulder

Assessments that are intended for instructional uses by teachers have extended histories in the field of literacy. The informal reading inventory has been refined by succeeding generations in the form of the running record (Clay, 1985) and miscue analysis (Goodman, 1968), but the underlying concept that teachers sample students' oral reading and retellings and study the strategies apparent in these responses has not changed substantially since Gray introduced the notion in 1920. As Johnston (1984) observed in the first *Handbook of Research on Reading*, informal reading inventories were first used at about the same point in time as the first generation of standardized silent reading tests. Educators and policymakers could have taken one of two routes 70 years ago. The choice of standardized tests and the relative obscurity of informal reading inventories fit the prevailing perspectives of literacy and accountability.

This emphasis (and many would say, over-emphasis) on standardized tests has repeatedly been called into question in recent years as perspectives on teachers, literacy, and assessment have changed. Calls for new assessments that integrate to a greater extent the involvement of teachers and information from classrooms have been heard in both the literacy (Goodman, Goodman, & Hood, 1989) and evaluation (Linn, Baker, & Dunbar, 1991) communities. To date, efforts to create alternatives have resulted in new forms of tests that use long texts from sources that children typically read (e.g., chapters from books,

---

[3] This paper was presented at the annual meeting of the American Educational Research Association, Atlanta, GA, , April, 1993.

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

magazine articles), open-ended responses, and questions that encourage thinking (see, e.g., Kapinus, Collier, & Kruglanski, in press; Weiss, in press). While these efforts to date have not integrated classroom-based assessments, they recognize the need for multiple indicators, where information that originates with teachers is used to tell part of the overall picture (Resnick & Resnick, 1989).

While literacy specialists have been involved in the design of these new assessments, many advocates of classroom-based assessments in literacy concentrate on the instructional applications, not the accountability uses (Cambourne & Turbill, 1989; Goodman et al., 1989; Johnston, 1989). Two perspectives can be seen in this literature (Valencia, Hiebert, & Afflerbach, in press). The first places the identification and use of assessments solely with teachers (e.g., Hansen, in press). According to this view, when teachers are given the responsibility for assessment, they will assess what is appropriate, when it is appropriate.

A similar emphasis on instructional uses of classroom-based assessments underlies the second type of project but a common core of assessments is encouraged in a school or district. Teachers adapt these measures for their contexts, but the assumption is that common goals of an educational community, like fluency or response to literature, can be captured by common classroom-based assessments like running records or literature logs (see, e.g., Au, in press; Gipps, 1993; Valencia & Place, in press). While the aim of these projects is to inform instruction, a common core of assessments also means that at least some of the information gained from the assessment can go beyond the individual teacher's classroom. Progress on a school's critical goals can be followed by teachers as children move across the grades. Since classroom-based assessments can capture children's efforts on goals and tasks that may be difficult to capture on large-scale assessments, classroom-based assessments have the potential to give educational agencies a broader perspective on student progress toward critical literacy goals.

Studies on teachers' use of assessment have primarily been of the first scenario of classroom-based assessment—where teachers are free to identify assessments. These studies suggest that the typical mechanisms for teacher change result in infrequent use of assessments by teachers. The teachers who were surveyed by Harris and Lalik (1987) used informal reading inventories

infrequently, even though considerable time had been devoted to their use in university coursework. Hiebert, Hutchison, and Raines (1991) found discrepancies across two case study teachers in a school where portfolio assessment had been identified as a focus. One teacher used informal reading inventories, studied students' writing samples, and took notes from children's conferences; the other teacher had little information to share with parents, children, or the research team. She claimed to store information in the "mental filing cabinet." In an extensive examination of teachers' implementations of assessments, Aschbacher (1993) found that teachers were more interested in the activities that could be represented in a portfolio than in the goals represented by these activities. Aschbacher concluded that this focus on activities rather than underlying goals created a barrier to sound classroom-based assessments.

Aschbacher's study (1993), like those of Hiebert et al. (1991) and Harris and Lalik (1987), summarized the manifestations of classroom-based assessment *after* coursework or staff development on classroom-based assessment. The present project was an examination of the issues that arise over the course of staff development on classroom-based assessments. Teachers grapple with many issues whenever they confront their practices, especially when those practices are interwoven with testing and accountability (Anders & Richardson, 1992). An understanding of progressions and issues should be informative to those in schools, districts, and other educational agencies like Chapter 1 who are contemplating initiating such projects.

Any set of issues or progressions reflects a particular set of structures and contexts. Two features of the particular contexts in this project should be noted. The first is the role of standardized tests. Teachers' concern and attention to their own assessments would be expected to be impacted by the stakes placed on standardized tests. During the year of this project, a moratorium had been called on standardized testing for the third grade in these schools. Consequently, standardized tests did not loom as an issue with these teachers. There was evidence, however, that the participating schools did not regard standardized tests as high stakes. This perspective had some validity since there was no evidence of high demands for test scores from either district administrators or from the state department of education, which has neither curriculum mandates or a high-stakes testing program.

Second, the progressions and dilemmas need to be viewed as a function of the particular workshop content. The project involved a collaboration between teams of teachers at school sites and a group of researchers. Teacher decision making was part of this effort. The intent of the researchers, however, was not to simply verify teachers' existing practices or to discuss possible uses of assessments as has been the manner in the first type of project described by Valencia et al. (in press). The university-based members of this collaboration began the project with the second perspective on classroom-based assessments. That is, the intent was to work with teachers within a school on a common set of assessments that captured key goals of the school and district. It was expected that teachers would adapt assessments to their classroom contexts. Further, since there were no school, district, or state mandates, teachers had the option of implementing these assessments. The underlying perspective of the research team, however, was that a school would identify some common assessments that captured shared goals. The progression of the workshops was influenced by this perspective in that initial sessions were devoted to the identification of shared goals and appropriate common assessments to capture those goals, and subsequent sessions involved developing scoring rubrics among participating teachers.

Method

Participants

An invitation to collaborate in a classroom-based assessment project was extended to third-grade teams of teachers in a school district that serves the northern region of a metropolitan area. Third grade was chosen because that is the grade at which state and district assessments most frequently occur (Linn, Graue, & Sanders, 1990). A requirement was that all third-grade teachers within a school would agree to participate. Three schools within the same district responded to the invitation and were chosen to participate. Because analyses by socioeconomic class would make it possible to identify individual schools (and jeopardize promises of confidentiality), such an analysis will not be conducted in this paper. However, the three schools represented a range of socioeconomic status, with 62% of the students eligible for the free or reduced lunch in one school and 3% and 9% eligible in the other two schools.

Data Sources

The design of the overall project extends beyond the workshop sessions that are the focus of the current analysis. Accomplishments on a performance assessment are being documented, with baseline and control group comparisons, for all students in classrooms of participating teachers and on teachers' assessments for a representative group of students. Interviews with teachers at three points in the year and transcripts of parent-teacher conferences will be available as well.

Data from transcripts of meetings over September to December of a school year provide the focus of this analysis. While meetings with teachers have extended over the spring semester, these have emphasized assessments with reading of informational text. Meetings over the fall semester were aimed at classroom-based assessments with the reading of narrative text. For two of the schools, seven meetings on literacy assessment occurred during this time period and, in the third school—Walnut (a pseudonym, which is the case for names of all schools and teachers—six meetings.

Each session was audiotaped and transcripts were made of these recordings. The fieldnotes of a research assistant were used to augment transcripts on features like the material that served as a reference to a teacher's comments and nonverbal cues. Two additional sources—the transcript of a May meeting prior to the fall implementation that was intended for teachers' sharing of their existing assessments, and applications that each school team submitted to the research team—were used to confirm patterns in teachers' views of literacy and assessment.

Foci of the Workshops

The sessions were designed to implement a long-standing perspective on curriculum-instruction-assessment (Cronbach, 1960) that has been adapted for classroom-based assessment (Calfee & Hiebert, 1991). Table 1 presents this overall structure.

This framework calls for identification of goals as the first step in the assessment process. Once goals have been identified, appropriate tools that represent those goals are identified. After data have been gathered using those tools, attention turns to the scoring and interpretation of results. In the case of

Table 1

Foci of Meetings with Teachers

Session	Content of meeting
April, 1992:	School team of Grade 3 teachers submit application for participation.
May, 1992:	All three school teams meet with research team for debriefing of performance assessment and to share their current assessments.
Session 1:	Shared presentation to all three schools on district goals and possible tools to assess those goals. Teachers meet as school teams to discuss goals.
Session 2:	Research team meets with each school individually to particular goal(s) and tool(s).
Sessions 3 to 7	Research team meets with schools individually on a bi-weekly basis to discuss through the implementation and use of the assessments. Session 6 included information from teacher-parent conferences.

classroom-based assessments, interpretation would lead to consideration of instructional practices.

In line with this framework, the first session began with the selection of goals. The research staff had studied the curriculum framework for the district. The framework reflected the whole language and writing process activity that had occurred in the district over the past seven years. Since the schools came from the same district and the district had a clearly written curriculum framework, all three school teams met together during the first half of the first session. The literacy specialist of the research team gave an overview of the primary goals of the district. Suggestions as to assessment tools/instructional activities that capture these goals were also given. Four goals were emphasized, each represented by a tool/instructional activity: meaning-making—summaries; fluency—running records; interest and extended reading—literature logs; self-assessment—annotations. These four tools had been identified by teachers as representing the goals of a literature-based or whole language program in a previous classroom-based assessment project (Valencia & Place, in press). The research team gave each teacher a notebook in which each goal was described and samples of classroom-based assessments for the tool for that goal were given. For example, after a description of the goal of meaning-making, samples of student summaries from texts that were chosen by students or teachers and teachers' scoring

rubrics were provided. After this overview, teachers spent the second half of the workshop meeting as school teams. Their task was to identify the goal(s) for which they would design and implement assessments over the fall semester.

At the next session which occurred the following week, the research team met with teams of teachers at their school sites. This session, for which teachers had several hours of release time, was devoted to identifying particular tools that would capture the chosen goal(s). All subsequent sessions were held on a bi-weekly basis with a team of teachers from a school. The agreement with the teachers was that these sessions would be an hour in length. Most sessions lasted longer than that, typically an hour and a half. These bi-weekly sessions were adapted to the concerns and issues of particular school teams within the framework already described (i.e., scoring and interpretation/instructional implications, following the identification of goals and tools).

Results

Steps of Analysis

There were three steps to the analysis. First, a set of categories was established. Transcripts that represented the three schools and three periods of time—first two sessions, middle three, and final two—were read and notes written down about the content of teachers' comments by two members of the research team. The two individuals would evaluate two transcripts and meet to discuss the categories that they had identified. They would return to another set of transcripts, using the common categories. The category scheme was refined until all of the interactions in the transcripts could be accounted for by the category scheme. This category scheme, which appears in Table 2, was used to code the data for all of the transcripts.

Next, all of the transcripts were coded using this category scheme. Themes were coded as conversations, with the number of turns that teachers contributed to a theme identified. The rounds of conversations for particular themes were clustered together to establish the weight given to a particular conversation during a workshop. Typically, a session had a primary theme (e.g., identification of goals), with a variety of sub-themes related to the

Table 2

Categories for Themes and Sub-Themes

I. GOALS
- A. Defining literacy goals
- B. Relationship to district/state goals
- C. Match of goals and tools

II. VIEWS OF ASSESSMENTS
- A. Standardized tests
- B. Performance/alternative assessments

III. LOGISTICS
- A. Setting up (Preparing materials—e.g., student-selected, same across classrooms)
- B. Execution (Frequency, unit)
- C. Scoring

IV. INSTRUCTION/ASSESSMENT RELATIONS
- A. Assessment event embedded or separate from instruction
- B. Instruction to support strategies being assessed

V. INTERPRETATION
- A. Interpreting student processes/performances based on data
- B. Interpreting based on perceptions of students
- C. Sharing interpretations—parents, children, other teachers.

overarching theme (e.g., how did fluency fit in with meaning-making). The primary themes, with their sub-themes, and secondary themes were established for each session.

Finally, these themes were examined as to the presence of a dilemma. A dilemma was defined as an obstacle, perceived or real, in teachers' implementation of assessments. An obstacle implies the presence of at least two divergent points of view—one where the issue is seen to be an obstacle and an alternative perspective where it is not seen as an obstacle (or at least a different obstacle). Often, the alternative point of view was held by the researchers. In some cases (and, as the analyses will show, in one school in particular), dilemmas were apparent among teachers within the school.

The analysis indicated that a focus on dilemmas provided only a partial picture of the interactions in and across sessions. Many of the sessions were consumed with discussions among teachers and/or the research team about a theme like how to characterize thorough or solid summaries. Such conversations involved questions and issues but the level of concern would by no means indicate a dilemma. The topics of sessions with no dilemmas are as informative as the presence of a major dilemma. Consequently, the decision was made to develop narratives about the sessions of each school team. These narratives describe the sequence of themes and, when they arose, dilemmas. The narratives also include descriptions of the views of goals and classroom-based assessments held by a school team as a whole. While no attempt has been made to analyze these narratives according to their underlying story grammars, the views of goals and assessments held by a team of teachers can be thought of as the setting and initiating event for the ensuing plot.

Descriptions of Dilemmas and Progression of Themes Within the Three Schools

Pine School. As was true in all the participating schools, the workshop sessions at Pine School proceeded along a particular topical path: establishing goals, determining tools for assessing the goals, implementing the assessments, scoring the products, and interpreting the results. But, of the three schools, the teachers at Pine moved the most rapidly from topic to topic. During a typical workshop, teachers quickly picked up project leader suggestions, worked through tasks, and generally came to some kind of closure at the end of each workshop. Few issues really grew into dilemmas and rarely resurfaced from one session to another. Pine teachers seemed intent on the techniques that project leaders suggested in sessions.

In the first session, teachers voiced a strong commitment to their district framework. One teacher referred frequently to the large binder of district literacy goals as the group deliberated on their project goal: "So are our choices like from 3.1, 3.3 [designations for three major literacy goals], or are they the little sections between them?" (Karen, 9/15, l.13-15). The conversation about goals moved to the conversation that would dominate the remainder of the session: tools for measuring student progress toward the goal. At the end of the session, Pine school's project goal was articulated. Despite the attention

to district goals, the teachers described their goal in terms of an instructional technique: "the nine week goal, then, is the story frame and the ending goal would be the 9 to 12 sentence narration" (Sara, 9/15, l.1089-1091).

The emphasis on techniques characterized what the teachers had done in assessment previously. They had used the Gates-MacGinitie (voluntarily since tests were not given until the end of third grade) for placing children into reading groups. While these previous assessment activities were mentioned during the second workshop, the teachers seemed intent on picking up another set of techniques—the techniques related to the running record and written summaries of text suggested by the literacy expert on the project team. They came to the third workshop with assessments completed on most of their students, ready to begin developing scoring schemes. They became the first of the three schools to describe instructional decisions based on the assessments and to describe instructional activities that would foster the processes captured by the running records and summaries. At the third session, Pine teachers described the use of the new assessments—running records—to change the reading materials for a group of students: "We thought that some of these kids could move up because we think Heath is easier than the ratings of the Gates-McGinitie. So where we have kids placed maybe in the 1.2, we're going to move them into the 2.1" (Elly, 9/29, l.5-9). This conversation surrounding the regroupings then shifted to instructional activities to support the literacy strategies being assessed. Discussion of instructional activities became the focus of much give and take among the teachers with one asking, "Now, so could you finish telling me, what will you do with Sue? She reads it. She blows it. Tell me the follow-up. OK. She blows it. Well, tell me, because I'm not sure what I have to do yet" (Elly, 9/22, l.593-596). Pine teachers speculated on various ways to connect the information from the assessments to instructional activities and agreed to continue addressing this issue in subsequent meetings, "But this is a beginning point, because if we do it at the end, hopefully because of this instruction, they will have improved" (Elly, 9/29 l. 550-552).

These first three sessions moved at a brisk pace at Pine. The next meeting took a different turn. The entire fourth session was devoted to the teachers' concerns about the project. The lead teacher for Pine school had brought a list of their concerns for both the mathematics and reading components of the

assessment project. The primary issue for the teachers, with regard to the literacy component, quickly developed into an impasse, a dilemma: how did the assessments fit with their instructional programs? The nature of this dilemma was summed up by the statement of one teacher: "The classroom comes first and extras come later and lately there have been lots of extras" (Lena, 10/13, p. 3 of field notes for workshop). The usefulness of the running records and summaries that had been described two weeks previously for shifting the level of children's reading materials was not mentioned. The teachers perceived that the assessments were taking time from instruction and from after school time that should be devoted to instructional planning.

The activity of the next workshop was designed to respond to this dilemma by focusing on the incorporation of running records into everyday tasks. Pine teachers returned to a focus on the task. The theme of "time" was brought up as a tangent of two conversations by the lead teacher of the team in this subsequent meeting, but was not pursued by any of the other teachers again, and did not resurface in any future meetings to the level of the fourth session. Rather, the nature of changes in instructional practices was a prominent theme in this session. Pine teachers described lessons about summaries and deliberated on appropriate ways to instruct and assess summaries. Several issues arose in this conversation. Teachers spent some time discussing sharing expectations with students, a point raised in an earlier session: "The fact is that kids before they start need to know the criteria. And we lots of times don't do that. I don't" (Elly, 9/29. l.635-637). One teacher raised the issue of using written summaries to assess reading comprehension by describing her scoring scheme then questioning the "two grades that we need to look at for a summary." This teacher went on to observe that "they can have all the information in the story, so you know they have the comprehension which is the main goal, but I picked--I thought we picked summaries at grade level last year because of the writing quality too" (Lena, 10/27, l.728-738). The issue of relying on writing to assess reading comprehension was one of the few issues to resurface in later sessions.

The separation or combining of reading and writing led to another conversation about how to instruct students in the writing of summaries. Instruction practices for the Pine teachers focused mainly on getting students to identify and record story elements. One teacher warned the group about

becoming too formulaic: "I think . . . modeling of how they could improve. Because my summaries right now are 'The characters are,' 'The story takes place,' 'The problem, the solution, what happened in the beginning, what happened in the end.' It's so rote, grocery list. It doesn't flow" (Lena, 10/27, l.1923-1928). As the teachers described their instructional activities they begin to grapple with establishing a consistent scoring rubric.

Discussions about instruction continued to dominate the final two sessions of the semester where teachers attended to instructional activities that fostered summarizing and, to a lesser degree, activities fostered by the running records. Large portions of these sessions were devoted to teachers' articulation of criteria for summaries identified as showing "thorough, solid, some understanding, or little understanding." Pine teachers' apparent commitment to a particular type of summary elicited concern about student engagement from one teacher in an earlier session, "I've found I can do it (summaries) only every 2 or 3 weeks because they don't like it" (Lena, 10/13, p. 5 of field notes for workshop). This concern reappeared briefly in this session and led to further sharing of instructional activities. One such activity involved students in scoring summaries themselves. As Lena explained, "So that's when I started writing terrible summaries and then they would tell me what was wrong with my summaries. But I think that there was some growth because the first week we started they didn't even know what a summary was and if I would have written a crummy one they couldn't have told (me)" (Lena, 11/17, l. 649-654). In discussing and developing a scoring rubric, teachers explored the issue of standards in scoring: "So the reason you choose a, you know, like on the way you're teaching it, you choose 'thorough' because it had all the elements and it was written the way you knew a summary should be written, not because that was one of the best ones out of the class" (Karen, 11/17, l.478-482). Again the conversation returned to instructional activities that can be facilitated by sharing the scoring rubric with the students.

Beginning with their identification of the story frame as the semester's goal, Pine teachers indicated an emphasis on techniques or the specifics of literacy, instruction, and assessment. This interest in specifics meant that they moved quickly in implementing assessments and in discussing the implications of the information gained from these assessments for instruction, such as grouping. But gathering running records on some students and

designating events for gathering students' responses to text and then scoring them required Pine teachers to rethink some aspects of their instruction. While this school had been involved in gathering children's work in portfolios, this team of teachers had been using a strand of assessments that occurred as separate events in their classrooms—the Gates-MacGinitie and a commercial informal reading inventory. These teachers had returned to a textbook series for reading instruction, after an extended period in the school where trade books had been exclusively used. The notion of embedding assessments within instructional events was one with which they grappled. The dilemma that these teachers had with the notion that assessments could be embedded in ongoing instructional events came to a head in the middle session. This session seemed to be profitable for both the teachers and the researchers in clarifying the underlying issues and moving to resolution. At the end of the semester, teachers were engaged actively in a conversation as to how instructional practices could support students' responses to text, especially summarizing. In the final session, the teacher who had continued to raise the dilemma of additional demands created by the assessments was able to articulate the discreteness with which they had viewed the assessments early on in the semester: "I'm sure when we pulled those [pages for running records], we kind of said, we're going to do this for the project so we're going to, throw something together. I mean that's honestly the way we did it" (Elly, 12/8, l.1234-1238). She also was able to reflect on the development that she had seen in students and to recognize her own learning: "It's been a good learning experience for me because I think this has been the best summary writing I've ever had with my kids" (Elly, 12/8, l.1915-1919).

Spruce School. During the first session, the definition of literacy that Spruce teachers agreed on for their semester emphasis was: "The first part is comprehension and not worry about the level and the variety of text. But just whatever they are reading, whatever group they are in, they are understanding" (Tamara, 9/15/92, l.240-244). Comprehension was the emphasis but teachers expressed some reservations about the forms that these responses can take. One teacher stated, "I think a lot of them right now if you told them to summarize a story they wouldn't have a clue what that meant" (Libby, 9/15/92, l.796-800).

This teacher's statement did not mean a reservation about written responses to text, however. Of the three schools, this school had had the most involvement with portfolios previously, with students conducting parent conferences with their portfolios in the spring. The teachers demonstrated this background and interest in portfolio assessment when they brought letters that they have had their students write about books that they were reading to the second workshop. The research team engaged the teachers in extending these assessments with a shared set of texts in a class or across the classes in the school. Teachers agreed with the idea of some designation of a core set of selections, since they were having difficulty scoring students' letters, which have been based on individual books. Spruce teachers identified selections in a core group of trade books that covered a range of difficulty from books that one or more of them intended to use over the semester.

By the third session, the majority of teachers had gathered written summaries or oral summaries and running records for most of their students. In line with teachers' concerns with an over-emphasis on fluency, the research team guided teachers in scoring the running records for semantic acceptability (the proportion of sentences that are meaningful). The session was spent productively, with teachers scoring their running records and reflecting on the oral and written summaries of their students. Teachers identified the need to select another book between *Little Bear* (a primer-level text) to *The Lemonade Stand* (upper second grade) since some children had read the former well but struggled with the latter. During this working session on scoring and identifying books, one teacher questioned the usefulness of this information for instructional purposes: "I can have different reading groups and stuff like that but they're, you know, it's not just reading, it's just universally, kids that just work on a very low level on all there is" (Tamara, 10/6/92, l.1438-1442). This teacher's comment did not consume a primary conversation in this workshop, however, as teachers worked on establishing students' proficiency levels and discussed the text difficulty of the books that they have had students read and summarize.

The differences in views of text difficulty were the source of a dilemma that dominated the fourth workshop. According to one teacher, the assessments went against the school's philosophy: "Our philosophy is kind of to look at the child and help them get from where they are to where they can

get during that year" (Libby, 10/20/92, l.229-232). The assessments had created conflicts for teachers. First of all, the assessments made some students aware that they weren't reading well as indicated in the statement: "I would say she's probably second grade now. But I hate labeling kids, and I hate making them feel bad" (Libby, 10/20/92, l.232-237). Libby went on to identify another problem with this assessment information—sharing of information with parents: "I would be real nervous about using this 'this is where your child is, this is a first-grade level.' . . . I don't want some parent going home thinking 'well, you've got to read to me every night this long because you are only reading at first-grade level' " (Libby, 10/20/92). Libby was the primary spokesperson for these concerns related to interpretation, but other teachers periodically chimed in with "I agree" (Reba, Penny). Tamara identified problems and possible benefits, stating the dilemma quite clearly: "And you know, sooner or later, some of them need to know where our kids are, and some of them need to know that they are reading at pre-primer at third grade. . . . On the other hand, it can be really bad on their egos" (Tamara, 10/20/92, l.367-375). The session ended with no real resolution.

When the research team met with the teachers again, the teachers who had been the quietest during the previous session reported on an instructional response to the dilemma. The assessments had led Penny and Reba to adapt their instructional practices by initiating whole-class phonics lessons: "We decided that we'd do our vowels. . . . So we did start with short 'a' " (Penny, 11/2/92, l.29-32). Libby raised questions about the appropriateness of this instructional practice for everyone: "Some of my kids are reading fifth-grade books with no problem at all. I mean, they can read just about anything that's put in front of them. Why would I want to do anything with short vowels with them?" (Libby, 11/2/92, l.408-414). Libby also described the instruction that she was providing some students: "All I'm saying is the strategies and the things we're talking about in the word patterns, those are the things that I'm doing with my, with two of my groups, but with my other one, we're just doing mostly character discussion and the plots of the story. And, you know, I'm not doing much skills with them and maybe I need to be doing something with them more advanced" (Libby, 11/2/92, 493-503). After an extended conversation that included comments like Libby's about the need for differentiation of instruction based on accurate information about students' proficiencies, Penny

equivocated about instruction of short vowels when she looked at her summary sheets of student performances on the running records and written summaries: "Now these guys, this one, she can read anything I give her. So they don't really need that much, just a little bit" (Penny, 11/2/92, l.639-642).

These statements suggest that the assessments had been the impetus for teachers' reflecting and acting on their instruction. They were still grappling with many issues surrounding text difficulty and student proficiency but gathering information on students' performances in their typical instructional tasks had led to changes in instruction. Libby carried the theme of differentiated instruction into the next workshop where she described the benefits of the changes in instruction for the students who are struggling readers: "But I really feel like when I have my groups, I feel like my, my most beneficial group is my low readers 'cause that's when I feel like the most teaching goes on is when I have them in a group" (Libby, 12/1/92, l. 1966-1970). The instructional implications were still being worked out by individual teachers as Tamara's statement indicates: "And there's not time in a day, it doesn't matter if it's my management or it's not my management to get these kids that are as low as they are, up" (Tamara, 12/1/92, l. 2614-2617). During visits to the classrooms at this time, the research team also discovered that Reba who had been quiet during most workshop sessions had been conducting running records almost daily and had been working hard with a handful of students whom the initial assessments showed to be reading poorly.

As part of a conversation during this session, teachers raised the need for more guidance on scoring of summaries and the final session of the semester was devoted to scoring of summaries. Teachers had an extended discussion about their students' written summaries. Unlike the early sessions where the school's philosophy was described as "just whatever they are reading," the discussion indicates high standards for student work as evident in Libby's description of her quandary in giving a high score to papers that are short:

> I think it is real solid too. It is hard sometimes when I read these because some of them are short and yet they seem to have all the characters and the problems and the solutions and the elements that, you know, and there's others that, you know, go into real detailed explanations about—so it's hard to know, you know, this has all the elements so this is a four, but then someone that added more detail to it, you know, got just a four, too you know what I mean? (Libby, 12/15/92, l.411-420)

Spruce teachers began with a global perspective of literacy and of assessment of children's progress. The examination of samples of students' oral reading and written responses created dissonance among the teachers. This information was viewed as potentially harmful because it might stigmatize students with the teachers themselves and children's parents. The teachers worked on the assessments, however, and began to discuss changes in their instruction. Some attempted guidance in word-level strategies for the group of students who the assessments had shown were not proficient readers. There were discrepancies among the teachers about how this instruction should occur. However, teachers were examining their instruction from a vantage point other than a view of comprehension regardless of text difficulty. Issues continued to arise but, at the end of the semester, teachers were examining their instructional practices and standards.

Walnut School. The first session at Walnut began with a spirited discussion about goals, and this spirited discussion or "grand conversations," to use Eeds and Wells' (1989) term, characterized ensuing sessions. The grand conversations or debates among Walnut teachers began with their identification of critical goals for the semester's work. This discussion focused on the relative role of fluency and meaning in proficient reading. After hearing comments that emphasized fluency, Judy cautioned that "they were separating the meaning from being able to read fluently" (Judy, 9/15/92, l85-187). A turn or two later in the conversation, however, Abby repeated the view that she and another teacher, Janet, had been stating: "I just feel like I need to have the kids read and I need to talk to them about what they've read to me that's fluency . . . yes, we're working towards meaning but I don't think that's a goal" (Abby, 9/15/92, l.205-208). And, Jackie brought these differences into the open by saying "I'm hearing two different things" (Jackie, 9/15/92, l. 257-258).

The discussion got heated in the second session where Judy argued against the use of a commercial reading inventory, saying "I give the IRI and I think this is an interesting score, okay? And how does it apply to the book the child is actually reading?" (Judy, 9/22/92, l. 1649-1651). The teachers discussed and resolved the issue by identifying parts of trade books that one or more of them used in their classrooms.

The topics of the conversations changed over the sessions—from defining literacy to developing a rubric that discourages formulaic responses to

instructional strategies that foster student thinking. Two characteristics remained constant across the sessions: the form of the conversation and continual references to the underlying processes that the assessments and instructional activities were intended to further. Teachers continually expressed their perspectives and asked questions about the perspectives of their colleagues. As part of these discussions, teachers recognized problems—with their instruction, the assessments, their students' progress to date. The pattern is consistent with Janet's response during the third session: "I wonder if we spend some more time on what a quality summary, that this will get better. Not so much that his reading understanding gets better, but what he's able to express will get better" (Janet, 10/13/92, l.1036-1038). And they worked on finding solutions. During the fourth session, Abby's concerns with the questions on their baseline assessment illustrated this stance: "We haven't scored them yet either because, especially the written, we weren't at all happy with what we got in and my concern and Janet's is that those questions are not getting . . . I'm not getting the information that I can use to tell parents how their kids are doing" (Abby, 10/13/92, l.40-44).

In subsequent sessions, teachers used the identification and implementation of classroom-based assessments to inform their instruction and their understanding of students' strategies. Teachers expressed disappointment with students' progress to date, as Jackie's comments during workshop five indicated: "Well, I'm disappointed because I wanted to take their rough drafts of what they presented to me as a summary and to be able to say, as a class, let's determine a one, two, three, and four. And I don't even have enough to do that. I don't have one summary that I think is good" (Jackie, 10/27/92, l.223-227). In these discussions, teachers considered the nature of the task. Jackie acknowledged that the summaries in question pertained to an entire book and that this was "an overwhelming task for the majority of the kids" (Jackie, 10/27/92, l.78-79). As often happened at Walnut, another teacher jumped in to provide a different perspective on children and their responses: "Another problem, the fact that what we think is important and what the kids think is important is often totally different? They may latch onto some funny little thing that happened in a story that really isn't that important" (Judy, 10/27/92, l.266-269).

Because of a weather cancellation and a schoolwide project that pushed a session into January, the last session of the semester occurred for teachers during the third week of November. Even at this point, Walnut teachers saw differences in their standards for students as a result of their reflections on assessment as illustrated in Janet's comment: "When you tell a kid 'now tell me what you predict' and expect a nice paragraph is much different than seeing a sentence that says 'I predict . . . He will go home.' We're expecting a lot more" (Janet, 11/17/92, l.639-643).

The airing of different views continued as teachers grappled with how to instruct these strategies. Judy raised a view different than that of the other teachers: "I'm almost finding, and I'm not sure if it's true or not, but summarizing seems to be a maturational skill, and some kids have it and other kids don't" (Judy, 11/17/92, l.923-926). Jackie responded to this comment with the observation that "I don't think the kids have had the practice and experience" (11/17/92, l.953-954). Janet also noted that the summarizing that children were being asked to do differed markedly from the retellings that they have emphasized in the past, observing that "we're almost giving kids mixed messages . . . Here you tell me you want me to write everything I think, and now you tell me you don't" (Janet, 11/17/92, l.1029-1036). Janet had identified a critical distinction in the literature on responding to text—the difference between summarizing and retelling. This observation came about through the extended conversations that these teachers had.

If an analysis were done on the amount of description or attention to portfolios in schools' applications to the project, the teachers at Walnut School would be described as the school with the least background of the three on classroom-based assessment. Walnut School's application had a single, passing reference to portfolio use in their school. Unlike Pine, they did not describe prior schoolwide staff development on workshops on portfolios or, unlike Spruce, implementation of student-led parent conferences with portfolios. What their application emphasized was a commitment to the district framework and the prior work of the school in establishing shared goals. Further, for the research team, the progress of the teachers at Walnut seemed slow at times. The portfolios of their students had fewer written summaries or running records than those of the teachers at Pine at the end of the semester. However, the understandings of underlying processes that the

assessments were uncovering and the descriptions of student growth, even in mid-November, were rich. The workshops provided the context for these grand conversations, and these conversations had been acted upon in assessments and instructional activities in classrooms. While providing incentive for these grand conversations about student progress on particular goals, the source of these grand conversations lay in a schoolwide effort over the past year. Walnut teachers began the semester with a foundation. They spent the semester building a solid structure of instruction and assessment on that foundation of perspectives on literacy.

Discussion

The task of embedding assessments like running records and written summaries elicited quite different responses from teachers. School teams had distinct struggles and issues. At all of the schools, working on classroom-based assessments put demands on teachers. At Walnut School, teachers used the context of the workshops to analyze and debate all aspects of an assessment—the materials for running records and summaries, the definitions of scoring schemes, and ways to teach so that students could gain the underlying proficiencies. In the other two schools, the embedding of the new assessments into teachers' instructional programs was the source of at least some level of conflict. The workshops took different progressions in those two schools. Pine teachers, with their emphasis on specifics, moved quickly in implementing the assessments and in using the information. They used the assessments to develop specific instructional techniques and specific scoring schemes. But, in many ways, the new assessments were seen as add-ons and were not firmly embedded in an underlying foundation. At Spruce, a global perspective on literacy and its assessment created the most dissonance of any of the schools with the classroom-based assessments. The assessments showed Spruce teachers that some of their students were not able to read beyond a very rudimentary level. They wondered if they should share that information with parents or with the students. They worried that this information would force them to label students. They struggled with how to go about guiding students who weren't reading proficiently.

The underlying views of literacy and assessment that a school team articulated during the first two sessions shed light on the particular struggles

and issues that arose during the implementation. The broader issue of why definitions of literacy and assessment differed across school teams can be speculated upon. For example, schools have different histories and leadership that at a given point can determine the individuals who are hired. Despite the presence of local histories and issues, these findings highlight several issues for specialists, administrators, policymakers, and other groups as they consider the implementation of classroom-based assessments.

First, attention needs to be given to what is mandated. The negative effects of standardized tests on curriculum have been recognized (e.g., Hiebert & Corley, 1993). To some, the short passages, low-level questions, and the limited responses have been the culprits, not the emphasis on assessment. Change the assessment to tasks that mirror the critical processes of literacy, it has been argued (Resnick & Resnick, 1992), and change the instruction. It is a rare assessment project that begins with a recognition of the transformation of goals and instructional practices in a state or district (e.g., Wixson, Peters, Weber, & Roeber, 1987). The teachers at Walnut who engaged in extended conversations to make the assessments work for them were the only team that had spent time on a schoolwide effort to define goals. The other teams of teachers began with a focus on the assessment techniques—in one case specific (story frames) and in the other global (student self-assessment like student-led conferences).

The project was designed to be an "assessment project" and, early on, there was an emphasis on assessments. But these assessments were emphasized because of the goals of literacy that they represented. Running records can be used in ways that capture a variety of processes. At one level, running records can be used to provide information on fluency only. At a broader level, however, running records provide information on meaning-making (through children's self-corrections and the semantic acceptability of their miscues). Written responses to text, too, can be used to establish a variety of strategies and goals. A perspective on literacy that can be described as constructivist or sociocognitive (Langer, 1991) was the basis for the selection of these measures. But the perspective itself was not the focus of separate sessions. Rather, sessions focused on assessment techniques. At Pine, the emphasis on assessments meshed with their technique orientation. It may well be that a focus on assessment techniques is a necessary first stage that at

least some teachers need to move through to a richer interpretation of literacy, student accomplishments, and assessments.

For the teachers at Spruce in particular, the underlying perspective on literacy—and the notion that text difficulty enters into the reading process—was discrepant with their global orientation. A three-month period is not sufficient to determine whether the implementation of assessments can change teachers' views of literacy and of literacy instruction, but there were indications that Spruce teachers were examining their views about their students' learning and their instruction.

Efforts that attempt to mandate classroom-based assessments should be implemented with an awareness that the responses of teachers may vary greatly. It should also be kept in mind that the teachers in these projects came from the schools that chose to participate in the project (with some indications of "nudging" from one or two of the principals). The willingness of teachers in schools that could not be nudged can only be speculated upon. Staff development specialists and state and district administrators who are looking at classroom-based assessments should be aware that the process is not necessarily one welcomed by teachers. Research efforts like those of the New Standards project will presumably be able to verify the extensiveness and nature of change between the first wave of volunteers and the follow-up cohorts when the assessments are mandated.

Ultimately, the issue of how children's learning is impacted by changes in practices, whether those emanate from new assessments or instruction, needs to be addressed. Future analyses in this project will permit examination of the ongoing portfolios, teachers' use of information with parents, and accomplishments on a performance assessment. The impact of participating in this assessment project on teachers' work with future cohorts of students also needs consideration. In particular, the nature of teachers' implementations when standardized tests are again administered in these schools should be studied.

The present analysis indicates that the collective view of a school on literacy and assessment manifests itself in different issues and dilemmas over the course of a classroom-based assessment project. As many have argued and demonstrated, regardless of the dimension of instructional or assessment

practices (see, e.g., Anders & Richardson, 1992), making changes in classrooms is a long and complex process. Initiating a shared set of classroom-based assessments in schools can contribute to the reflection and debate that underlies this process of change.

References

Anders, P., & Richardson, V. (1992). Teacher as game-show host, bookkeeper, or judge? Challenges, contradictions, and consequences of accountability. *Teachers College Record, 94*, 382-396.

Aschbacher, P. (1993). *Issues in innovative assessment for classroom practice: Barriers and facilitators* (CSE Tech. Rep. No. 359). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Au, K.H. (in press). Portfolio assessment: Experiences at the Kamehameha Elementary Education Program. In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities*. Newark, DE: International Reading Association.

Calfee, R. & Hiebert, E.H. (1991). Classroom assessment of reading. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 281-309). New York: Longman.

Cambourne, B., & Turbill, J. (1990). Assessment in whole language classrooms: Theory into practice. *The Elementary School Journal, 90*, 337-349.

Clay, M.M. (1985). *The early detection of reading difficulties* (3rd ed.). Portsmouth, NH: Heinemann.

Cronbach, L.J. (1960). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English, 23*, 4-29.

Gipps, C. (1993, April). *Emerging models of teacher assessment in the classroom*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Goodman, K.S. (1968). The psycholinguistic nature of the reading process. In K.S. Goodman (Ed.), *The psycholinguistic nature of the reading process* (pp. 13-26). Detroit: Wayne State University Press.

Goodman, K., Goodman, Y., & Hood, W. (Eds.) (1989). *The whole language evaluation book*. Portsmouth, NH: Heinemann Educational Books.

Gray, W.S. (1920). The value of informal tests of reading achievement. *Journal of Educational Research*, 103-111.

Hansen, J. (in press). Literacy portfolios: Windows on potential. In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities*. Newark, DE: International Reading Association.

Harris, L.A., & Lalik, R.M. (1987). Teachers' use of informal reading inventories: An example of school constraints. *The Reading Teacher, 40*, 624-630.

Hiebert, E.H., & Corley, R. (February, 1993). *A comparison of students' reading on standardized and performance assessments in a high-stakes testing context.* Submitted for publication.

Hiebert, E.H., Hutchison, T.A., & Raines, P.A. (1991). Alternative literacy assessments: Teachers' actions and parents' reactions. In S. McCormick & J. Zutell (Eds.), *Learner factors/teacher factors: Issues in literacy research and instruction* (40th Yearbook of the National Reading Conference, pp. 97-105). Chicago, IL: NRC.

Johnston, P.H. (1984). Assessment in reading. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1). White Plains, NY: Longman.

Johnston, P.H. (1989). Constructive evaluation and the improvement of teaching and learning. *Teachers College Record, 90*, 535-549.

Kapinus, B.A., Collier, G.V., & Kruglanski, H. (in press). Maryland School Performance Assessment Program: A new view of assessment In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities*. Newark, DE: International Reading Association.

Langer, J. A. (1991). Literacy and schooling: A sociocognitive perspective. In E.H. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 9-27). New York: Teachers College Press.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 15-21.

Linn, R.L., Graue, M.E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "Everyone is above average." *Educational Measurement, 9*, 5-14.

Resnick, L.B., & Resnick, D.L. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer.

Valencia, S.W., Hiebert, E.H., & Afflerbach, P. (in press). Realizing the possibilities of authentic assessment: Current trends and future issues. In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities.* Newark, DE: International Reading Association.

Valencia, S.W., & Place, N. (in press). Literacy portfolios for teaching, learning, and accountability: The Bellevue Literacy Assessment Project. In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities.* Newark, DE: International Reading Association.

Weiss, B. (in press). California's New English-Language Arts Assessment. In S.W. Valencia, E.H. Hiebert, & P. Afflerbach (Eds.), *Authentic reading assessment: Practices and possibilities.* Newark, DE: International Reading Association.

Wixson, K.K., Peters, C.W., Weber, E.M., & Roeber, E.D. (1987). New directions in statewide reading assessment. *The Reading Teacher, 40,* 749-754.

TEACHERS' IDEAS AND PRACTICES ABOUT ASSESSMENT AND INSTRUCTION[4, 5]

Hilda Borko, Maurene Flory, and Kate Cumbo

CRESST/University of Colorado at Boulder

This paper presents results of preliminary analyses of interviews with teachers about their knowledge, beliefs, and practices related to assessment and instruction in reading and mathematics. In particular, we describe and examine teachers' knowledge and beliefs about what it means to know reading and mathematics, and their ideas and practices related to assessment and grading in the two subject areas. We also examine the nature and extent of changes in these characteristics during the first semester of a year-long intervention designed to help the teachers develop performance assessments in reading and mathematics.

The literature is replete with examples of how difficult it is for teachers to change their knowledge, beliefs and practices in fundamental ways (e.g., McDiarmid, 1990; Putnam, Heaton, Prawat, & Remillard, 1992; for a review, see Borko & Putnam, in press). Yet, many current educational reform movements are calling for just such fundamental changes. The performance assessment movement is no exception. Teachers are being asked to reject traditional forms of classroom assessment, such as the objective, paper-and-pencil unit tests and chapter tests typically found in mathematics and reading

[4] This paper was presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 1993.

This research is part of a larger research project, Studies in Improving Classroom Teaching and Local Assessments, and was supported in part under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

All members of the research team contributed to the conceptualization and data collection for this paper

[5] We extend special thanks to the teachers who are participating in the project. Without their support, commitment, and countless hours spent in workshops and interviews with us, this project would not have been possible.

textbooks, and to replace them with performance assessments. The claim is that these more "authentic" performance assessments, with their focus on problem solving and higher-order thinking skills, will improve teaching and enhance student learning (Wiggins, 1989; Resnick & Resnick, 1992).

For teachers to incorporate such fundamental changes into their ongoing instructional practices, they must have the knowledge necessary to implement the changes and the beliefs to support them. Given the literature on teacher change in other areas, it seems reasonable to expect that such transformations in teachers' ideas and practices about assessment might be difficult to accomplish. This expectation receives some support from members of the assessment community such as Stiggins (1991), who warns that many teachers lack the "assessment literacy" necessary to implement performance assessments.

Our research team's commitment to classroom-based performance assessments, coupled with our concerns about teachers' ability to implement such assessments in their classrooms without support and guidance, were prime motivators for the larger project of which this study is a part. The specific commitment of this paper's authors to understanding and improving the process of learning to teach led to this particular study of changes in the ideas and practices of the teachers participating in the project.

Methods and Data Sources

This study is one component of a 3-year research project. The major purposes of the larger project are to work with third-grade teachers to develop classroom performance assessments in reading and mathematics, and to study the effects of performance assessments on teachers' knowledge and beliefs, instructional practices, and student learning. The project uses a "bottom-up" intervention, in which the research team is meeting with teachers on a weekly basis to provide support and guidance in the development of classroom performance assessments. Our intention in these weekly workshops is to foster changes in the teachers' assessment practices by helping them to think about instructional goals, develop or select assessment tasks appropriate to those goals, and articulate scoring criteria.

Participants

Primary participants in the project are the third-grade teachers at three elementary schools in a school district on the outskirts of Denver. Schools wishing to be considered for the project submitted a short proposal and documentation of approval by the principal, all third-grade teachers, and the parent accountability committee. Participating schools were selected to represent a range of socioeconomic levels and student achievement.

Data Sources

Data sources for this paper are interviews with the third-grade teachers conducted before we began our weekly workshops (September 1992) and again at the end of the Fall semester (January 1993). Semi-structured questions focused on teachers' knowledge, beliefs and practices related to assessment, instruction, and the relationship between assessment and instruction. Interviews were conducted by members of the research team, at the participating schools, during the school day. All interviews were audiotaped and transcribed.

Data Analysis

Our analyses for this paper examined most closely teachers' responses to a subset of interview questions that addressed their conceptions of what it means to know mathematics and reading, their reports of assessment and grading practices, and their beliefs about the advantages and limitations of performance assessments. We also considered participants' responses to other interview questions, when they were relevant to the issues that were the focus of this paper.

We developed a coding system based on our research questions and an initial reading of a sample of Fall interview transcripts. The system included categories such as "know-math," "how assess-math," "why assess-math," "grade-math," and their parallels for the subject area of reading. We added several categories that focused on change (e.g., "change know-math," "change assess-math") after reading a sample of the Winter interviews.

Our analysis consisted of coding the data separately for each school, identifying major themes in the data, writing cases of the individual schools organized around these themes, and then conducting a cross-case analysis to

identify similarities and differences in the patterns of knowledge, beliefs, and practices that characterized teachers at the three schools. This paper presents the cross-case analysis. We conclude with a discussion of the nature and extent of change in the teachers' ideas and practices, and with some hypotheses about factors that may explain patterns in our findings.

Patterns in Teachers' Knowledge, Beliefs, and Practices

Our cross-case analysis begins with teachers' knowledge, beliefs, and practices related to reading, and then examines parallel issues for the subject area of mathematics. For the most part, we focus on school-level patterns and trends. Whenever individual teachers provide particularly important insights into the change process, we discuss them as well.

Reading Programs

At a very general level, teachers in the three schools held a shared conception of what it means to know reading. The two major dimensions that comprised their conceptions were skill and understanding. Affect played a consistent, but much less prominent, part in their responses.

Teachers at all three schools identified fluency and word-level reading strategies as important components of the skill dimension at the beginning of the school year. Abby's (all names used in reference to the project are pseudonyms) description of one characteristic of excellent readers captures their views: excellent readers "would know what to do when they came to a word that they didn't know. They'd have the skill to figure that out; they'd use all sorts of strategies . . . that we, as readers, use and don't even think about."

The ability to identify story elements was seen as a central component of "understanding" by teachers at all three schools. At a more fine-grained level of analysis, however, conceptions of this component of reading differed across the schools. Teachers at Walnut wanted students to "identify story elements—character, setting, events, problems, inference based on actions or personalities of characters." Teachers at Spruce talked in vaguer terms about their desire that students be able to "make sense" out of a story by identifying story elements. At Pine, teachers talked about identification of story elements as if it were a skill to be mastered in a fairly routinized manner. To facilitate mastery, they had developed a sequence of structured tasks to help students

move from responding to a series of questions asking for specific words or phrases in the text that identified each story element, to writing a paragraph in which story elements were reported.

Other components of reading for understanding that were mentioned by teachers differed across schools. At Walnut, the ability to tell others about what one has read was a prominent element of understanding. Teachers at Pine talked about reading for meaning beyond the literal aspects of the text. Teachers at Spruce seemed to consider word-level meanings, in addition to "making sense" out of a story, as a component of understanding.

At a general level, Fall interviews also revealed similarities in the assessment and grading practices at the three schools. Teachers at all three schools assessed for both skills and understanding, and they all used paper-and-pencil tasks as well as observation and listening approaches.

Again, differences emerged when a more fine-grained analysis was conducted. These differences were associated with one basic difference in the schools' reading programs. Reading at Walnut and Spruce was organized around a Reader's Workshop approach and, for the most part, children read in self-selected trade books. In accord with Reader's Workshop, a major portion of assessment consisted of individual conferences with students about the books they were reading. All teachers at the two schools had developed systems for keeping written records of the conferences.

For teachers at Spruce, assessment consisted primarily of "listening," "observing," and "looking at student writing." The teachers seemed to place considerable emphasis on fluency and word-level understanding as they listened to students and asked questions about their reading. All of the teachers kept written records of conferences, although there were individual differences in the extensiveness of these records.

Teachers at Walnut were much more explicit in describing both what they looked for in their conferences and what information they recorded. As Abby explained, referring to how she determined what her students know and can do in reading, "Most of it is by listening to them read, and talking to them, conferencing with them, asking direct questions . . . that give you a feel for what the kids are keying into, or if they can key into character development or anything like that." For Jackie, conferences also provided information about

the affective component of reading: "Are they enjoying the book? . . . Do they like it?" Although their specific record keeping systems differed, the teachers seemed to be keeping track of similar kinds of information. For example, Jackie kept anecdotal records in a notebook, with information such as the book a child is reading, his or her reaction to that book, oral reading proficiency, and areas in need of instruction. "In the past, I have had a notebook that has tabs, one for each child. So that I have one notebook, and everybody's name is there, and there's notebook paper in there, and basically it's just keeping an anecdotal record...kind of tracking what the kids are doing." She tried to keep these records "while I'm conferencing. . . . that's the best time to do it, because it's fresh in your mind, you have 26 students you're working with, and you just can't remember which student needs work in oral expression, and which student needs help in self-selection, and where the kids are."

In contrast to both schools, Pine teachers used a basal textbook series and placed students in ability groups. For assessment purposes, these teachers placed more emphasis on traditional paper-and-pencil tasks and written work than did their colleagues at Walnut and Spruce. Although they acknowledged the importance of anecdotal records, only half of them had devised systems for keeping track of these records, and none seemed comfortable with any strategy they knew of for taking the records into account when determining grades. The overriding characteristic of their view of reading assessment was its problematic nature. As Elly noted, "I think reading is extremely hard to evaluate."

Despite these differences across the schools, a key theme among all teachers was their desire to learn more, through work on the CU project, about keeping informal records during the reading program.

Changes in reading programs. Teachers at all three schools had incorporated two types of activities into their reading programs—running records and written summaries—by the time of the Winter interviews. Both of these activities were introduced by the CU project's literacy expert at workshops early in the year, as potential assessment tools. Perhaps the most consistent change associated with participation in the CU project, in either reading or mathematics, was that all teachers were using running records (although to varying degrees) with their below-grade-level readers. Teachers talked about their use of running records in very similar ways. Most teachers

at each school reported that the use of running records increased their awareness of students' reading abilities. Many of the teachers reported that information they acquired from running records informed their instruction. Sara's comment captured both of these ideas: ". . . things like running records . . . tend to tell you what types of errors they are making, and they do change my instruction as to what I need to focus on with these kids." Perhaps one reason running records were adopted by the teachers is that, as one teacher explained, they are "a quick way to follow the kids' progress," and they "are much easier than any other kind of reading test I've found."

All of the teachers also reported that they were using written summaries as part of their reading assessments. As was the case with running records, they talked about written summaries, and apparently used them, in similar ways. All teachers developed scoring criteria for summaries (typically, in conjunction with their teammates, and in part during their bi-weekly workshops with CU project team members). They shared these criteria with their students and, to varying degrees, students participated in the evaluation of their own written summaries. Teachers at all three schools reported an increased awareness of their students' understanding of text. They also reported that students appeared to be more aware of their own reading ability and their teachers' evaluation criteria, and better able to evaluate their reading performance (i.e., their written summaries). As one teacher at Walnut explained, the students gained "an understanding of what teachers expected from them. . . . I don't think some kids ever had a clue as to how we got their grades in reading."

We did find several differences across schools in their discussions of running records and written summaries. In general, teachers at Walnut and Pine were more explicit than their colleagues at Spruce in their discussions of the specific kinds of information revealed by these new assessment tools. As one teacher at Walnut explained, running records "tell us specifically, maybe some sounds and letter blends that kids don't know. . . . They can also tell us which kids recognize their errors and do something about them."

Teachers at Walnut discussed student changes to a greater extent than did teachers at the other two schools. They addressed affective (e.g., feeling better about themselves as readers) as well as cognitive (e.g., use of known spelling patterns, understanding of teachers' expectations and grading

practices) dimensions of student change. For example, Janet noted about the below-grade-level readers in her class and their experiences with running records, "I think kids that are struggling can look at those [words that they read correctly] and say, 'look at all these words I know,' and kind of get away from 'I am a bad reader'."

Also, as a result of their experiences with running records and new insights about their students' reading abilities, teachers at Walnut decided to study the Benchmark reading program as a team, in order to learn specific word-level strategies to incorporate into their instructional programs. Judy talked about the differences that these experiences had made for her. "[I]n the Fall . . . I did not feel that I knew how to meet the needs of the first-grade readers. And I was running out of material for them to read. . . . And I didn't have the children that I had pushed beyond third-grade level, because I wasn't aware of what they were able to do." By the time of the Winter interview, she was "helping some of those kids who are having reading difficulty . . . by looking at the words that they are having difficulty with and picking out the small key spelling patterns and sounding them out that way."

Another difference between schools was that teachers at Walnut viewed using written summaries as a way of assessing reading competence to be problematic, to some extent. Their concern was that some students are proficient readers, but are not able to write good summaries. They were unwilling to penalize students' reading grades because of weaknesses in writing. As Janet explained, "The summaries, we took into consideration, but certainly did not use those as a major part of the reading grade because we have good readers who read well, who recognize nonsense, who understand stories and reading, who were not writing great summaries. And we did not feel that we could grade them low." The teachers at Spruce, in contrast, saw no problem in assessing reading proficiency with techniques that incorporate writing skills. All Spruce teachers seemed to view reading and writing as going "hand-in-hand." Two talked explicitly about the connection between reading and writing. Teachers at Pine did not address this issue.

Finally, one aspect of individual differences among teachers is important to note. For at least a small number of teachers, changes in their assessment and instructional practices seemed to be accompanied by changes in their conceptions of reading. Janet's comment represents such a shift: "We have

made an error in assuming that all kids learn to read from immersion in literature and building vocabulary. And that works for a lot of kids, but not for all." This insight led Janet to depart from some of the tenets of a Reader's Workshop approach to teaching reading, and to "incorporate a little more phonics" and explicit instruction in strategies such as compare/contrast and key spelling patterns in a school where "it seems like that for a few years, 'phonics' was a four-letter word." In contrast, other teachers saw similar changes in assessment and instructional practices as the addition of new strategies into their repertoires, but considered these additions to be consistent with their existing conceptions or philosophies.

Mathematics Programs

As was true with reading, at a general level, teachers' conceptions of what it means to know mathematics were similar across the three schools and included skills, understanding, and (to a lesser extent) affective components. Differences in emphasis across schools were greater in mathematics than in reading. For teachers at Pine and Spruce, the primary emphasis in instruction and assessment, at least for the first quarter, was on math facts and skills. All of the teachers at Pine stressed that "[g]ood mathematicians know their facts." A common goal for the first grading period is "Making sure that they are memorizing their math facts through 18."

Teachers at Walnut acknowledged that students must "master the facts," and that computation and number facts are components of "knowing mathematics," but they found the focus on skills to be problematic. Janet was among those most hesitant in attributing importance to mathematics skills. Although she acknowledged that "we do spend a little bit of time in the room working on mastering the facts" and she was not ready to say "that algorithms . . . shouldn't be taught," she was quick to point out that "the computation that we do is really a means to an end." That end, for her, was mathematical understanding and problem solving.

At all three schools, the meaning of "understanding mathematics" was ambiguous. In its fullest sense, understanding seems to include the ability to explain steps in a mathematical procedure (for example, the ability "to explain and describe how they derive an answer") as well as understanding the conceptual underpinnings of mathematical procedures and the connections

among concepts and procedures (for example "number sense," "recognizing patterns," or "seeing connections"). It typically was not possible to determine whether and to what extent individual teachers' conceptions included both components. This was less true at Walnut, where teachers' comments were more extensive, and where their conceptions appeared to incorporate both components. For example, when asked what it means to be excellent in mathematics, Abby responded " . . . that they have a good understanding of numbers and when you would use addition, when you would use subtraction. . . . They can see a connection between patterns and a connection between numbers. . . . They can explain how they came up with an answer, maybe even give more than one way to solve . . . "

Teachers' conceptions of problem solving were also ambiguous. To the extent that we can tell, however, they seemed to differ across the schools. Teachers at Walnut and Spruce did little more than mention problem solving as one aspect of what it means to know or be excellent in mathematics. For example, students who are excellent in mathematics can "use information from many sources to apply to the problem," and "know what resources to go get, to help [solve a problem]." Teachers at Pine were somewhat more explicit in their comments. Interestingly, more than their colleagues at the other schools, Pine teachers seemed to focus on problem solving as a set of procedures and steps that can be taught and learned in a fairly routinized manner. This orientation to problem solving was similar to their view of summarizing in the subject area of reading.

In mathematics, in contrast to reading, assessment strategies at all three schools focused on skills to a greater extent than understanding. All of the teachers were fairly dependent on traditional forms of assessment, particularly when it came to assigning grades in mathematics. And, none of the teachers had well-developed systems for keeping track of information gathered while observing and listening to students during math lessons.

Beyond these general patterns, however, school-level differences in assessment strategies were greater in math than in reading. Timed tests in the four basic operations are mandated by the school district and administered at all three schools. They were much more prominent at Spruce than at Pine or Walnut, both in the teachers' interview responses and in their mathematics programs. In fact, teachers at Spruce generally were more focused on skills

and facts than were their colleagues at the other schools. They were also much vaguer in their descriptions of other forms of assessment they use.

Teachers at Pine relied almost exclusively on other paper-and-pencil tasks to supplement timed fact tests for assessment and grading purposes. They were much more explicit than their colleagues at Walnut or Spruce, in describing the various types of tasks they use; their repertoire included daily written work (usually problems copied from the book), chapter tests from the books, and personally-created end-of-unit tests.

In contrast, observation and listening seemed to play a much more prominent role in the assessment systems of teachers at Walnut (though not in their grade determination). More than teachers at the other schools, they talked explicitly about what they look and listen for when observing students engaged in mathematical activities. Jackie was the most specific. She did "a lot of observing what they need to complete a task and how successful they are with that task. . . . those students who are frustrated, who don't understand and who perhaps are relying on a partner or relying on the teacher for procedures." One specific focus of her observations was students' use of manipulatives. She would "give them a task and give them the option of using manipulatives or not; it becomes quite apparent those students who do not need the manipulatives." She gave several examples of assessment for specific topics; for place value, addition, and subtraction: "I am watching them and observing how quickly they're able to set up a problem or work a problem with the beans, and then again describe what they're doing and how they're approaching it."

Changes in mathematics programs. In contrast to the subject area of reading, there were more school-level differences than similarities in the nature and extent of changes in mathematics assessment. There were, however, some consistencies across the schools. For example, the mathematics educator on the research team supplied the teachers with many suggestions for mathematics activities that use manipulatives and focus on higher-order thinking and problem solving. Teachers at all three schools incorporated these activities into their mathematics programs.

By the mid-year interviews, teachers at the three schools were asking students to provide written explanations for some of the math problems they

solved. Similarly to the written summaries they incorporated into the reading assessments, they developed scoring criteria for these explanations, taught the criteria to students, and had students participate in the scoring process. They also reported that they knew more about their students' understanding of mathematics than was the case when they relied exclusively on more traditional forms of assessment. Jackie's comment, though more explicit than most, represents the teachers' sentiment: "I think it [new types of assessment] gives us insight into the student's thinking and learning, more so than just a score or plus number. So I think what we have done with the project has really taught us as educators to question kids to think about how did you get that solution or what process did you follow. . . . I certainly learned . . . I think I know more about my students and where they're coming from than just are they right or wrong."

The schools differed, however, in the extent to which they used these new activities and explanations for assessment purposes. Teachers at Walnut clearly considered the activities and explanations to be part of their assessment systems. Each teacher developed a record-keeping system for the new activities, which included information about both written assignments and teacher observations. Janet used checks, plusses, and minuses. As she explained, the activities "lent themselves more to a check system" than to percentage scores or grades. Jackie kept track during observations with "a check sheet with everybody's name on it, and at the top, just the skill I'm looking for. And I would note whether they were showing mastery or needed more work." She used a 3-point numerical system to score written problem-solving assignments. Most teachers reported relying less on percentage scores on traditional paper-and-pencil tests, and more on these records of student performance, when evaluating students' mathematics proficiency and assigning report card grades.

The situation was more complex at Pine and Spruce. Teachers at these schools clearly considered the activities to be part of their mathematics programs. They talked about them, however, primarily as instructional tools. The activities were seen as important means for engaging students in higher-order thinking and getting them to "communicat[e] their answers in language." The teachers acknowledged that they were learning more about their students and their mathematical understanding by using these

activities. For example, "I think it tells me much more about the whole student as a learner, as a thinker. [Before] the only thing I looked at was if they knew their facts." However, most teachers continued to assess students with the traditional chapter tests and timed tests. Our sense is that, at the time of the Winter interviews, Pine and Spruce teachers viewed the new components of their math programs—with the exception of written explanations—as instructional activities which provided them with new insights about their students, rather than as assessment strategies *per se*.

Finally, as was the case in reading, we found some evidence that for at least a small number of teachers, the changes in their instructional and assessment practices seemed to be accompanied by fundamental changes in their conceptions of mathematics. Judy presents the clearest example, as reflected in her comment that "I'm looking at math differently." In the Fall, her response to the question about what it means to be excellent in mathematics mentioned, in general terms, aspects of understanding such as "manipulate numbers," "estimate," and "having a feeling for numbers . . . and what they meant." At mid-year, her answer was much more detailed, and she was more explicit about mathematical processes such as "explain how they got an answer," and "analyze her own mathematical thinking." She also expressed the belief that children can be excellent in mathematics because they "do things quickly and easily and accurately" or because they "ask great questions and [are] very thoughtful about problem solving and what [they do]. . . . So there are really two kinds of mathematicians."

Our sense is that Judy is an exception to the rule, and that most teachers did not experience fundamental changes in their conceptions of mathematics. As Janet explained, "I don't think that for me there has been a big philosophy change from last year to this year. . . . The project didn't give me a new philosophy. What it has done is push me on a little more in the philosophy I already had."

Speculations About Change

At this point in the project, it is premature to draw conclusions about the nature of changes in the teachers' knowledge, beliefs or practices that are resulting from their participation in the project. At most, we can offer some observations and tentative speculations about the future.

It is clear that in both reading and mathematics, teachers have added to their repertoires of activities. In reading they clearly consider these activities—running records and written summaries—to be new forms of assessment. Further, they reported new insights about students' reading, and they seem to have greater understandings, at least about below-grade-level readers' skills, based on these assessments. One thing that is less clear, and that appears to be less consistent across schools and teachers, is the extent to which teachers have changed their instructional programs to take advantage of these new assessments and insights.

In mathematics, teachers have consistently incorporated new activities into their instructional programs. However, only some of the teachers consider these activities to be assessment tools. Also, the extent to which the activities have provided the teachers with greater insights into their students' mathematical understandings is not clear from their interview responses.

What is perhaps most unclear to us at this point in the project is the extent to which teachers' conceptions of reading, learning to read, mathematics, and learning mathematics have changed as a result of their experiences on the project and the new tools they have incorporated into their programs. We have evidence that a few of the teachers have begun to reconsider their conceptions. Most of the teachers, however, seem to have incorporated new ideas from the project into their existing reading and math programs, without making fundamental changes in these programs or in the conceptions of reading and math that underlie them. These patterns are neither surprising nor discouraging to us, given what the literature suggests about difficulties in effecting teacher change. Further, given the degree of change we have seen in some schools and some individual teachers, we are optimistic that if the teachers continue to make changes in their programs, to reflect on the impact of these changes on student learning, and to share their experiences in the weekly workshops, we will see more extensive changes in their conceptions by the end-of-year interviews.

References

Borko, H., & Putnam, R. (in press). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. Guskey & M. Huberman (Eds), *New paradigms and practices in professional development*. New York: Teachers College Press.

McDiarmid, G.W. (1990). Challenging prospective teachers' beliefs during early field experience: A quixotic undertaking? *Journal of Teacher Education, 41*(3): 12-20.

Putnam, R., Heaton, R.M., Prawat, R.S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. Elementary School Journal, 93, 213-228.

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Stiggins, R. (1991,). Assessment literacy. *Phi Delta Kappan, 72*, 534-539.

Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan, 71*, 703-713.

PARENT OPINIONS ABOUT STANDARDIZED TESTS, TEACHER'S INFORMATION AND PERFORMANCE ASSESSMENTS[6]

Lorrie A. Shepard and Carribeth L. Bliem

CRESST/University of Colorado, Boulder

New forms of assessment are seen by many as the centerpiece of current educational reform efforts. A growing body of research has documented the negative effects of standardized testing on teaching and learning (Madaus, West, Harmon, Lomax, & Viator, 1992; Shepard, 1991; Smith, 1991). Traditional forms of testing—filling in the blanks and picking right answers—have contributed to an implicit national curriculum focused on low-level skills (Resnick & Resnick, 1992). At the same time, efforts to gain external control over classrooms via standardized measures and teacher-proof curricula have led to the deskilling and deprofessionalization of teachers (Darling-Hammond, 1988; McNeil, 1988). In contrast, reformers expect performance assessments to have a non-distorting and positive effect on instruction and learning by aiming at more direct, complete, and integrated measures of desired student outcomes.[7]

Although there is widespread enthusiasm among educators for portfolios and other forms of performance assessment, implementation is problematic for both technical and political reasons. In this paper we address one formidable obstacle—the objection of parents.

---

[6] This paper was presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April, 1993.

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

[7] In this paper we use the term performance assessment as defined by Resnick and Resnick (1992). "A performance assessment is one that uses direct judgments and evaluations of performances rather than indirect indicators of competence" (p. 61). We consider performance assessments to be synonymous with *direct* assessment as described by Frederiksen and Collins (1989) and *authentic* assessment described by Wiggins (1989).

Policymakers perceive a demand for standardized testing from both parents and the general public. Standardized test score results are deemed essential to monitor individual student progress and to hold schools accountable. This demand is evidenced by annual Gallup Poll results demonstrating the high percentage of respondents favoring the use of standardized national tests by public schools in their community to measure the academic achievement of students (Elam, Rose, & Gallup, 1992). As a result, legislators, school board members, and professional educators fear that parents will be "up in arms" if performance assessments are adopted to supplant familiar standardized tests. In the early stages of the current research project, for example, sympathetic members of the district accountability committee asked that researchers discontinue the use of the phrase "alternative assessment" because it connoted abandonment of standards and rigor.

The purpose of this study was to examine in greater depth parent opinions about standardized tests and performance assessments. The research was guided by the following questions. 1. How do parents in this sample respond to marker Gallup Poll questions about the quality of local schools, the quality of *information* provided to parents, the desirability of standardized national tests, and the potential uses for standardized test results? 2. How do parents evaluate the usefulness of standardized tests compared to less formal types of information such as report cards, talking to the teacher, or seeing samples of their child's work? Do parents differentiate between information they find useful in learning about their child's progress in school versus the information they use to make judgments about the school as a whole? 3. After being shown sample questions and tasks, what preferences do parents express regarding district use of standardized tests and performance assessments in reading and mathematics? Is it true that parents disdain the use of performance assessments as less rigorous or objective? What features of the two types of measures affect parent preferences? 4. How do parents respond when asked to make their judgments about standardized test questions and performance assessments in the context of classroom instruction?

Methods

Larger Research Project

The present study of parent opinions was conducted in the context of a larger research and development project. In the larger project researchers worked with third-grade teachers for the entire 1992-93 school year to develop classroom performance assessments in reading and mathematics. The central aims of the research were to examine the effects of implementing performance assessments on instruction and student learning. Papers by Borko, Cumbo and Flory (1993), Flexer and Gerstner (1993), and Hiebert and Davinroy (1993) consider changes in teachers' thinking about assessment and instruction using beginning-of-year and mid-year interview data and transcript analysis from weekly workshop sessions. Effects on student achievement will be assessed at the end of the year by comparing third graders in the three participating schools to those in three control schools on the Maryland Assessments for reading and mathematics. (The accuracy of this quasi-experimental comparison is strengthened by matching of schools on socio-economic variables, baseline comparisons of the two sets of schools on the Maryland Assessments in May 1992 before the start of the study, and covariate adjustments based on premeasures administered to third graders in September 1992.) A collateral study also addresses changes in students' perceptions of teacher evaluations during the course of the project.

Parent Questionnaire and Interview Samples

The study was conducted in a working-class and lower-middle-class school district on the outskirts of Denver, Colorado. The district was selected on the basis of three criteria: (a) an ethnically diverse student population, (b) a history of standardized accountability testing, and (c) a willingness on the part of district personnel to seek a 2-year waiver from standardized testing in the three schools selected to participate.

The three participating schools were selected because the principals and third-grade teams of teachers volunteered. As part of the formal process required by the Colorado Department of Education to apply for a waiver from standardized testing, approval was sought from each school's accountability committee. Thus, although data on teachers' beliefs and practices suggest

that teachers were fairly traditional in their approach to classroom assessment, the volunteer nature of the school samples makes it probable that the school communities in this study were slightly more favorably disposed toward performance assessments than are schools in general.

Three control schools were also identified to be used for comparative purposes when analyzing student achievement, teacher beliefs, and parent opinion. The control and participating schools were matched on free and reduced lunch percents, percent minority, and other knowledge of neighborhood similarities. The following display illustrates socio-economic differences among the three participating schools as well as their matches to control schools.

	Participating Schools			Control Schools		
	1	2	3	1	2	3
Free and Reduced Lunch	61%	9%	6%	55%	13%	3%
Percent Minority	37%	16%	14%	45%	19%	10%

Questionnaire parent samples were selected in both the participating and control schools. Interview parent samples were identified only in the participating schools. The sampling design was developed such that mutually exclusive but randomly equivalent stratified samples of parents could be used to measure change in parent opinions over time. Data reported in this paper, however, represent only the pre-measures collected in the fall at the end of the first academic quarter. Questionnaires were sent to a random one-third of parents in each participating and control classroom. Overall, response rates were 70% for the participating schools and 49% for the control schools resulting in samples of 69 and 36 respectively.

In the participating schools, parents of three children per classroom were sampled to be interviewed in conjunction with first-quarter parent conferences. Children were stratified by achievement level (high, middle, and low). Teachers were provided with two alternate names within each stratum to use in cases where parents declined to participate or where scheduling precluded participation. We have no accurate way to estimate what biases were introduced into the sample by self-selection or schedule conflicts.

However, it was a relatively rare occurrence for teachers to pick the second name on the list because a conference was going to be problematic or because a parent was thought to be unresponsive to school requests. Ironically, a greater percentage of parents were willing to commit to the time demands of the interview than returned questionnaires. After the fall interview sample was agreed upon (n = 13 classrooms x 3 = 39), attrition occurred in the sample because of absence from the conference (n = 5) and technical failures in recording (n = 1) resulting in a final interview sample of 33 parents or parent dyads. Absence from the scheduled parent conferences was related to student achievement level; 3 of the absent parents had children in the low ability group and 2 were from the middle group. Therefore, the final sample was not as balanced as originally intended.

Questionnaire and Interview Instruments

The four-page mailed questionnaire had three parts. The first part asked parents to rate the usefulness of report cards, talking to the teacher, standardized tests, and seeing graded samples of their child's work. The second part provided displays showing typical questions on third-grade standardized achievement tests and typical performance assessment tasks in both reading and math. Parents were then asked to rate their degree of approval for each of the four types of measures. Then they were asked to show their preference for one or the other type of measure when used in instruction. The third set of questions was taken from the 1992 Gallup Poll. Parents responded to the questionnaires anonymously.

Interviews, conducted immediately following parent conferences, were guided by a structured but open protocol. The interview was divided into three parts, two of which closely paralleled the parent questionnaires. Parents were first asked broad questions aimed at eliciting their conceptions of reading and mathematics knowledge and what they used as indicators to judge their child's progress in each area. Two forms of the interview were printed so that half of the interviews began with reading questions and half began with mathematics. The second set of questions asked parents to rate the usefulness of report cards, talking to the teacher, standardized tests, and seeing graded samples of work. Parents were asked to think aloud and elaborate on their answers while filling out the rating form. For the last set of questions, parents

were shown more extended examples of questions from standardized tests and performance assessments in reading and mathematics. They were asked about their approval or disapproval of each and about the advantages or disadvantages they saw with each type of measure. A final question asked parents which of the two kinds of tests they would like to see used in classroom instruction.

Interviews were audio recorded and transcribed. Extensive field notes were also taken. In one case where a parent declined to be recorded, answers were typed directly into a computer.

Findings

Gallup Poll Results

Nationally 71% of the general public as well as 71% of public school parents favor requiring public schools in their community to use standardized national tests (Elam et al., 1992). If the purpose of the present study is to go beyond this superficial endorsement of standardized tests and develop a deeper understanding of parent opinions, then it is important to know how similar parents in the study sample are to the national population of parents.

Table 1 provides a summary of questionnaire results for third-grade parents in the study compared to data from the 1992 Gallup Poll (Elam et al., 1992). A preliminary analysis showed no significant differences between questionnaire responses for control and participating schools, therefore they are combined here as a single sample.

On the most classic Gallup question (#21) concerning national tests, there is a disconcertingly large disparity between third-grade parents in this study and the national sample of public school parents. Only 46% of local third-grade parents favor national tests compared to 71% nationally. Opposition rates are more similar because a very large proportion of third-grade parents (25%) responded in the Don't-Know category. On other questions, the gap between the two groups is much narrower suggesting that the third-grade sample can be thought of as only slightly more negatively disposed toward standardized testing than the nation as a whole. On questions 22-27 addressing various uses of standardized national tests, the percentage of third-grade parents endorsing each use was consistently 12% to 14% less than

Table 1

A Comparison of National Results and Third-Grade Parent Responses (n=105) on 1992 Gallup Poll Questions

18. **Students are often given the grades A, B, C, D, and Fail to denote the quality of their work. Suppose the public schools themselves, in this community, were graded in the same way. What grade would you give the public schools here—A, B, C, D, or FAIL?**

	A	B	C	D	FAIL	Don't Know
1992 National %	9	31	33	12	5	10
3rd-Grade Parent Sample %	13	53	25	2	2	5

19. **How about the public schools in the nation as a whole? What grade would you give the public schools nationally—A, B, C, D, or FAIL?**

	A	B	C	D	FAIL	Don't Know
1992 National %	2	16	48	18	4	12
Public School Parents National %	2	17	48	18	4	11
3rd-Grade Parent Sample %	1	8	50	20	7	15

20. **What grade would you give the information you receive from your third grade child's teacher(s) regarding his or her academic progress—A, B, C, D, or FAIL?**

	A	B	C	D	FAIL	Don't Know
1992 National %	31	30	23	7	6	3
Public School Parents National %	42	29	20	6	2	1
3rd-Grade Parent Sample %	32	37	20	6	1	4

21. **Would you favor or oppose requiring the public schools in this community to use standardized national tests to measure the academic achievement of students?**

	1992 National %	Public School Parents National %	3rd-Grade Parent Sample
Favor	71	71	46
Oppose	20	23	30
Don't Know	9	6	25

In addition to measuring the academic achievement of students, do you think these standardized national tests should be used or should not be used for the following purposes?

	1992 National %			3rd-Grade Parent Sample		
	Should	Should Not	Don't Know	Should	Should Not	Don't Know
22. **rank public schools**	65	26	9	53	37	10
23. **grade promotion**	60	32	8	46	45	9
24. **teacher pay**	38	52	10	26	62	12
25. **school funding**	36	54	10	22	68	10
26. **improve instruction**	79	14	7	72	17	11
27. **student weaknesses**	85	9	6	73	19	8

for the national group, with the exception of using test results to improve teaching where the difference was only 7%.

On the more general Gallup questions dealing with quality of schools and quality of information provided by schools about each child's academic progress, the third-grade parents rated their schools higher (66% gave As and Bs compared to 40% for public school parents nationally); they rated the nation's schools lower (9% As and Bs vs. 19% nationally); and they rated the quality of information they receive just the same as did a national sample of elementary parents (69% vs. 71% As and Bs).

Given the above similarities and dissimilarities between national samples of parents and the third-grade sample here, we suggest that the findings in this study—about broad parent preferences and especially about reasons parents give for their preferences—are likely to be generalizable beyond the confines of this particular district. However, it is unlikely that strict quantifications, such as the percent preferring performance assessments over standardized tests, can be generalized either to nonrespondents in the same district or to the nation as a whole.

Parent Opinions About Standardized Tests and Informal Sources of Information: Questionnaire Results

On the questionnaire, parents were asked to evaluate the usefulness of four types of information for two different purposes, learning about their child's academic progress and judging the quality of the school. Data in Table 2 show the percentage of responses for the first purpose. Overwhelmingly, parents believe that they learn the most about their child's progress by listening to the teacher; 77% rated this source of information as very useful. A total of 94% rated hearing from the teacher in the top two categories. All three informal sources of information about student progress—report cards, talking to the teacher, and work samples—received strikingly higher usefulness ratings than did standardized tests. However, it should be noted that a biasing factor in this study is that these third-grade parents were surveyed in the fall before the first opportunity for their third-grade child to be tested. Therefore, parents were likely to be responding on the basis of experience with older children or a general opinion they held regarding standardized tests. Parents who had little knowledge of

Table 2

Parent Questionnaire Ratings of the Usefulness of Each Type of Information for Learning About *Their Child's Progress in School* (n=105)

Place a check in the appropriate column to indicate the usefulness of each type of information.	Not at all useful 1	2	3	4	Very useful 5	Blank/ Missing
1. **Report cards**	2%	2%	20%	33%	43%	
2. **My child's teacher talking about his/her progress**	0	2%	4%	17%	77%	
3. **Standardized tests**	6%	15%	41%	22%	14%	2%
4. **Seeing graded samples of my child's work**	0	0	10%	30%	60%	

5. **Pick one type of information you think is most helpful in learning about <u>your child</u> and tell what you learn from it and why it is helpful.**

Report cards (n=5)

"I think the most important item is report cards because it shows how well they are doing in each subject and then we can determine where help is needed."

Teacher talking (n=44)

"Talking with my child's teacher is most helpful because I learn firsthand what progress is being made in class, where the shortcomings are and how I can best help at home."

Standardized tests (n=0)

Graded samples (n=38)

"This way I can see the actual work, the teacher's response, and evaluate what I understand the child's level of learning to be at."

"When we are able to see our child's work we can see the progress that he makes and can help him in the areas that he needs help."

Other (n=9)

standardized tests and whose oldest child was in third grade may not have had sufficient information to judge the usefulness of such tests.

In Table 3 parent ratings are presented for the same four types of information but for the purpose of judging the quality of the school. As

expected, more parents considered standardized tests to be useful for this purpose than for learning about the progress of individual children (45% vs. 36%). In addition, 19 respondents (18%) wrote in comments suggesting that it is the comparative information provided by such tests that makes them useful for evaluating the school's performance.

However, much of the data in Table 3 are contrary to the expectation that parents would prefer external and formal measures for school accountability purposes. Indeed the greatest number of parents, 84% and 83% respectively, valued "my child's teacher talking about his/her progress" and "seeing graded samples of my child's work" to help them evaluate the how good a job the school was doing. Furthermore, in follow-up comments parents gave justifications that showed they were mindful of the change in purpose from the first set of questions to the second. Parents say that these informal sources help them learn about the quality of education by giving them first-hand information about the school curriculum, teacher knowledge and expectations, school policies, and classroom climate.

Given earlier differences found between the study sample and the national population on Gallup questions, we were concerned that the above, somewhat surprising findings—downplaying the usefulness of standardized tests—might be true only for parents in this sample and not generalize to a broader population of parents. Therefore the present sample was subdivided on the basis of Gallup Poll question 21 into those parents who "Favor" use of standardized national tests and those who "Oppose" (parents answering Don't Know were omitted). Predictably these two groups differed in their ratings of the usefulness of standardized tests on Question 3 in Table 2. However, they did not differ at all in their much higher ratings of informal sources of information. For example, on Question 2, teacher talking, the mean ratings were 4.65 and 4.74 for the Favor and Oppose groups respectively. For Question 4, seeing graded samples of work, the means were 4.58 and 4.45. (Neither difference is statistically significant and the slight numeric differences are offsetting.) On the second set of questions dealing with accountability purposes, the two groups again differed significantly in their usefulness ratings for standardized tests ($\bar{X}$ Favor = 3.94; $\bar{X}$ Oppose = 2.80). They did not differ in their ratings on the other three sources of information; both groups gave report cards low marks for this purpose but rated teacher

Table 3

Parent Questionnaire Ratings of the Usefulness of Each Type of Information to Evaluate the *Quality of Education Provided at Their Child's School* (n=105)

Place a check in the appropriate column to indicate the usefulness of each type of information.	Not at all useful 1	2	3	4	Very useful 5	Blank/ Missing
6. Report cards	18%	11%	35%	17%	16%	3%
7. My child's teacher talking about his/her progress	4%	2%	18%	28%	46%	3%
8. Standardized tests	8%	14%	30%	20%	25%	4%
9. Seeing graded samples of my child's work	1%	4%	20%	27%	46%	3%

10. Pick one type of information you think is most helpful in learning about the quality of education provided at your child's school and tell what you learn from it and why it is helpful.

Report cards (n=0)

Teacher talking (n=26)

"Talking with the teach allows you into the school environment to measure emphasis, attitudes, some insight into treatment of students (i.e., respect, encouragement, etc.)."

"Again I would say talking with the teacher in person. That way I get an understanding of the school policies and programs and what curriculum the school uses."

Standardized tests (n=19)

"Standardized tests show me where [my school] ranks in comparison with district, state and national averages —"

Graded samples (n=18)

"Again, samples of my child's work show me what it is they are doing in the classroom, and will show me where the school's emphasis is at."

Other (n=25)

"Newsletters from school sent home with the student. These discuss what the goals are that have been established, what type of work the students are doing, what teachers expect, etc."

"Being able to be in the classroom, getting to know the teacher, seeing hands-on, spending time and helping at school."

talk ($\bar{X}$ Favor = 4.02; $\bar{X}$ Oppose = 4.10) and work samples ($\bar{X}$ Favor = 4.26; $\bar{X}$ Oppose = 4.00) highly. This means that the group favorable to standardized testing, which may represent a greater proportion of the national population, only considered standardized tests to be as useful as talking to the teacher for accountability purposes and not as useful as seeing graded work samples. For the purpose of learning about an individual child's progress, both groups agree strongly that all three sources of information are more useful than standardized tests.

Parent Opinions About Standardized Tests and Informal Sources of Information: Interview Results

In the interview data, parents reported even more glowing opinions about the usefulness of informal types of information for learning about their child's academic progress than they had on the questionnaire. In Table 4a and Table 5, quantitative results are presented from the interview questions corresponding to the questionnaire data shown in Table 2. The question about standardized tests was pulled out and treated in a separate Table to allow for separate analysis of the qualitative data. Although the parent responses about the usefulness of standardized tests are virtually the same (36% of the questionnaire sample said they were useful vs. 33% in the interview sample), the interview respondents were even more positive about the usefulness of

Table 4a

Parent Interview Ratings Regarding the Usefulness of Informal Types of Information for Learning About Their Child's Progress in School (n=33)

How useful is each type of information listed below in helping you learn about your child's progress in school?

	Not very useful				Very useful	Blank/
	1	2	3	4	5	Missing
a. Report cards	0	3%	18%	15%	61%	3%
b. Talking with the teacher	0	0	0	6%	91%	3%
d. Seeing graded samples of my child's work	0	0	0	15%	82%	3%

Table 5

Parent Interview Responses Regarding the Usefulness of Standardized Tests for Learning About Their Child's Progress in School (n=33)

How useful is each type of information listed below in helping you learn about your child's progress in school?

	Not at all useful				Very useful	Blank/
	1	2	3	4	5	Missing
c. Standardized tests	12% (n=4)	21% (n=7)	27% (n=9)	21% (n=7)	12% (n=4)	6% (n=2)

Normative Information n=4 ($\bar{X}$ =4.7)

"It compares him to everybody else and lets you know how he's doing compared to them. Because if he's just getting As on everything and everybody else is getting As for not doing anything at all, then it doesn't say much for him."

"How they do compared to other children is very important so that you have an idea that they are progressing like they should."

Helpful Academically n=4 ($\bar{X}$ =4.2)

"I think they are very useful. Not to me but to the teachers. I think they probably should have these so the teachers can inform the parents as to where they're at. Informs me as to what she needs help on."

Other Positive Attributes n=3 ($\bar{X}$ =4.0)

Children as Individuals n=6 ($\bar{X}$ =3.0)

"Some children just do not respond well to certain standardized tests. I think that children are different types of learners and standardized testing sometimes does not always catch that, the full breadth of a child's learning abilities."

"A standardized test really doesn't center around what the problems are with the individual child, but are used as a measurement of how much this child knows."

Invalid Results n=6 ($\bar{X}$ =2.7)

"I think it's got a lot to do with how well a child takes tests, what their particular morning was like, or whether they had a good night sleep. I think there's so many variables that can influence their score on the standardized testing."

Parents' Own Experiences n=2 ($\bar{X}$ =1.5)

"I don't like them. And I probably have to say because when I was a kid they made me nervous. I do better without the test. And when they go to give you that big test, it's overwhelming and you know you're being graded on whether you're stupid or smart."

Don't Know n=2 ($\bar{X}$ =2.5)

Child hasn't taken test or Parents have never seen results n=3 ($\bar{X}$ =2.0)

Other negative attributes n=2 ($\bar{X}$ =1.0)

Unfamiliar with standardized tests n=1 (parents left question blank)

report cards, hearing the child's teacher talk about progress, and graded samples of work than the questionnaire respondents had been. For example, 77% had rated teacher talk as "Very Useful" on the questionnaire, whereas 91% of the parents interviewed gave the "Very Useful" rating. Because parents in the interview setting were asked to fill out a rating sheet identical to the question set used on the questionnaire, it is unlikely that the demand characteristics of the interview (having to explain their answers, etc.) account for the more extreme answers in the interview. The more likely explanation is that parents were interviewed immediately following parent conferences. It is also possible that the interview and questionnaire data differ because they are based on different samples; the interview data were collected only in participating schools as compared to the questionnaire data collected from both control and participating school parents. However, there were not significant differences in the questionnaire results between the two samples on these questions.

Transcript analyses of parent comments explaining their ratings are summarized in Table 4b and 5. In Table 5, parent comments about why standardized tests are useful or not useful are related predictably to their ratings (means are shown in parentheses). Parents who rated standardized tests highly commented on the usefulness of normative comparisons and the ability of tests to identify academic strengths and weaknesses. Parents who gave lower ratings to standardized tests tended to comment on the vagaries of test performance on any given day or their own negative experiences taking standardized tests.

Parents had a great deal to say about why they valued informal sources of information, especially talking to the teacher, to learn about their child's progress. Note that all three informal types of information, even report cards which got the lowest rating, were rated as more useful than standardized tests. Parents want to know in some detail how their child is doing. They want to know strengths and weaknesses and what they can work on at home. To interpret the meaning of where their child is, they seek comparative information, that is, how their child is doing in relation to grade level expectations or compared to the rest of the class.

Perhaps the most interesting finding, however, is that parents, by and large, trust their child's teacher to provide them with this type of information.

Table 4b

Parent Interview Responses Regarding the Usefulness of Informal Types of Information for Learning About Their Child's Progress in School (n=33)

Nearly three-quarters of respondents (73%, n=24) appreciate how these types of information provide feedback about how well their child is doing academically. They cite understanding their child's strengths and weaknesses as well as seeing evidence of his or her progress over time.

"If I didn't have the report card, there would be no way I'd be able to judge that he was actually being able to retain what he's learning. She (the teacher) shows me what I really need to work on with him and with him being there, he sees that he needs to work on this."

"The report cards are very useful. You can (see by the) quarterly report the improvement she's made or if she's not making improvement, you can tell."

"As long as she [the teacher] can see in math that he's improving with each of his tests, that's all she cares about. It's just as important that he is doing the best he can do, and that he's improving."

"I like (graded samples). I think it's useful because then I know from (the teacher) showing from September to now, I know that she's improving."

Over one-third of the respondents (36%, n=12) comment on the normative aspects of this information. They depend on the teacher to provide the comparative information.

"I think report cards are very useful because their pluses and minuses (system of grading) tell you where they're at in reference to the rest of the class basically."

"(Graded samples) give you more of a clear idea of where the child stands within the class itself."

"Well, (these types of information) let me know that they are doing their level of work. I know that they're at the level that they should be doing. That assures me that they are doing the level that they're supposed to be."

"Well, I think on the report cards, it's very useful because in her reading you know what level she is reading."

With regard to talking to the teacher, respondents value the opportunity for communication with the teacher (18%, n=6).

"If there's questions, just keeping communication with the teacher is important. If you have questions, sometimes things come home that you don't understand and you need to know the answers."

"Talking with the teacher, I think, is most important. We can then ask questions as to the report card ..."

Talking with the teacher helps 24% of the parents (n=8) because the teacher spends time with their child.

"They [teachers] see what work they're doing more than we do because we both work. They're with our kids 9 hours a day."

Table 4b (continued)

"She's the trained professional, she knows what to look for if something should come up that we should be aware of...To be able to talk with somebody who can see their development and be there all times is very important."
"She deals with my child every day. She knows my daughter and can tell me things about my daughter's learning that she notices. She sees her in a context where she's maybe one-on-one at times, she's in a small group, she's in a large group, and sees my daughter in a wide variety of experiences, where I see her only in a family setting."
"They can see everything that they're doing, and they can see if they're messing around, not getting their assignments, or if they're struggling to understand..."

Note. Data from the three questions in Table 4a were treated as an aggregate. The themes emerged from the parent comments. Categories are not mutually exclusive.

"She's the trained profession, she knows what to look for . . . " "She deals with my child every day. She knows my daughter and can tell me things about her learning." We considered the very high approval ratings given to teacher information to be striking because usually the public endorsement for standardized national tests is taken to mean that parents require external verification of progress and cannot rely on teacher judgment.

When the purpose shifts from "learning about my child's progress" to evaluating the school, the interview results parallel the questionnaire findings but again with more extremely positive ratings for informal sources of information from the interviewed parents. Quantitative ratings are shown in Tables 6a and 7. Just as with the questionnaire data, the usefulness of standardized tests increased when the purpose was to judge the school rather than an individual child's progress. On the questionnaire, the increase in positive ratings was from 36% to 45%; for the interview sample, the increase was from 33% to 45%. Again on the question about standardized tests, there was a close correspondence between the responses from the questionnaire and interview samples. In explaining their ratings, approximately one-third of the parents interviewed said something about the value of standardized tests for making comparisons. "That is one of the reasons I like the standardized tests, because to me if you have a national standard test for third graders, it shows you where your kid is against national standards which doesn't necessarily say anything about your kid but it might point out there is a problem here . . . "

Table 6a

Parent Interview Ratings Regarding the Usefulness of Informal Types of Information for Learning About the Quality of Education Provided at the School (n=33)

How useful is each type of information listed below in helping you learn about the quality of education provided at this school?

	Not at all useful				Very useful	Blank/
	1	2	3	4	5	Missing
a. **Report cards**	12% (n=4)	18% (n=6)	27% (n=9)	12% (n=4)	24% (n=8)	6% (n=2)
b. **Talking with the teacher**	6% (n=2)	3% (n=1)	12% (n=4)	6% (n=2)	70% (n=23)	3% (n=1)
d. **Seeing graded samples of my child's work**	6% (n=2)	3% (n=1)	9% (n=3)	24% (n=8)	55% (n=18)	3% (n=1)

Table 7

Parent Interview Responses Regarding the Usefulness of Standardized Tests for Learning About the Quality of Education Provided at the School (n=33)

How useful is each type of information listed below in helping you learn about the quality of education provided at this school?

	Not at all useful				Very useful	Blank/
	1	2	3	4	5	Missing
c. **Standardized tests**	18% (n=6)	6% (n=2)	24% (n=8)	21% (n=7)	24% (n=8)	6% (n=2)

Normative Information n=12 ($\bar{X}$=4.2)

"The standardized test would probably be very useful because they compare with other students in other schools."

"That is one of the reasons I like the standardized tests, because to me if you have a national standard test for third graders, it shows you where your kid is against national standards which doesn't necessarily say anything about your kid but it might point out there is a problem here..."

"The standardized testing, I guess, if it's compared to other school districts, it would have to be somewhat useful in understanding the quality of education."

"The standardized tests—there are so many things that fall into each district, I don't know how I feel about those. Yes, I think it would. Because of course the parents are going to want their child to go to the better schools."

Table 7 (continued)

Curriculum Information n=4 ($\bar{X}$ =4.0)

"It shows you what the kids are doing, what he does know, what he can remember, what he's learning in the class. It gives you an idea about what they're presenting to the kids."

"I would feel that that would be very useful with the school. That they're following certain guidelines.

Other n=9 ($\bar{X}$ =2.9)

"As parents, we just go by what we have here. We don't make a comparison as far as what's going on elsewhere. We just kind of take it for granted what's here."

One parent replied that the standardized tests aren't useful because her child has learning disabilities and has a hard time reading the test.

Child hasn't taken test or parents haven't seen results n=3 (x=2.3)

Test not important n=2 ($\bar{X}$ =2.0)

"I just don't feel it's that important."

Test provides no information n=2 ($\bar{X}$ =1.0)

"I don't think that shows any effect on the school even."

"That doesn't tell me anything about the quality of education here at this particular school. It just tells me that these are tests that pretty much had to be given out during the end quarter of the semester."

Unfamiliar with standardized tests n=1 (parents left question blank)

When rating informal sources of information, the results from the interviews can be seen as a more extreme version of the pattern found in the questionnaire data (as occurred for the first set of questions on children's progress). Even for the purpose of evaluating the school, parents find talking to the teacher and seeing graded samples of work to be more useful than standardized tests. In Table 6b categories of parent responses are presented with illustrative quotations. In support of their positive ratings, parents explained that talking with the teacher allowed them to see what expectations were being set, the quality of the curriculum, and how caring the teacher is with students. Specifically parents said that seeing the actual work that students brought home let them judge whether what was being taught was worthwhile.

Table 6b

Parent Interview Responses Regarding the Usefulness of Informal Types of Information for Learning About the Quality of Education at the School (n=33)

Talking with the teacher helps identify her expectations and goals (n=4)

"Talking with them gives you a better feel of the goals that they have in mind for the kids, the criteria, teacher criteria, the goals for different grade levels. Usually when we talk about their progress it's in relationship to those kind of things which is an indication to me of the quality of the education here."

Talking with the teacher is telling about her personality and style with students (n=7)

"I think if there's a lot of kids that are having problems, that maybe it reflects the teacher. You know, so to me, that's where I would say how it affects the school."

"I think that's very useful, very significant to tell you that the teachers are really caring, or that they're really good with the students."

"Talking with the teacher tells me a great deal about the quality of education because she is able to tell me about herself, you know, her strengths, her weaknesses, and things like that."

Seeing graded samples (and to a lesser extent, talking with teacher) allows an evaluation of the curriculum (n=15)

"That (graded samples) lets me know whether my child's learning and I think the paperwork that they get lets me know what they're doing."

"I think (graded samples are) very useful. I like to know what kind of things because not everything that they teach, you think is useful."

"I just think if it's a really good cross section of what they do, the samples, then it would be a pretty good indication of the quality. If it's just random cases of their work, or ones that they just save because they like and they put them in the portfolio, or something, then I don't think it's a real good indication."

"I can see what kind of work the teacher is handing out. The teacher is the one that's in there quarterbacking the classroom. You know, if she's handing out pretty basic stuff to the kids to work on, then that's pretty boring, you know, "get-me-through-the-school-day" type of activities as far as I'm concerned. But if she's handing out stuff that will keep their interest and get their initiative going as far as keeping them active in school, and wanting to learn, that pretty much sets the tone for the school year and gives me an indication of what kind of quality teachers there are, and what kind of quality programs are here at this school."

Report Cards are not useful because they're about the child (n=3)

"That's a 2 because that doesn't tell anything about the school, that tells about the student."

Report Cards are not useful because of subjective grading (n=4)

"(They) wouldn't be very useful in telling the quality of education because you're not really sure what they are being graded on until it has been explained to you or you have see other things."

"...because everybody grades differently."

Report Cards are not useful—Other reasons (n=10)

> I can see what kind of work the teacher is handing out. The teacher is the one that's in there quarter-backing the classroom. You know, if she's handing out pretty basic stuff to the kids to work on, then that's pretty boring, you know, "get-me-through-the-school-day" type of activities as far as I'm concerned. But if she's handing out stuff that will keep their interest and get their initiative going as far as keeping them active in school, and wanting to learn, that pretty much sets the tone for the school year and gives me an indication of what kind of quality teachers there are, and what kind of quality programs are here at this school.

The only consistently negative comments regarding informal measures pertained to report cards. About half of the sample said that report cards were not very useful for judging the school because "they don't tell about the school" or because grading criteria are ambiguous or noncomparable across schools.

Parent Reactions to Standardized Tests and Performance Assessments: Questionnaire Results

On the questionnaire, parents were provided with displays showing typical multiple-choice questions that appear on traditional standardized achievement tests and a sample of more open-ended questions used in performance assessments. Separate displays were shown for both reading and math. Although more examples were given on the questionnaire, the following are illustrative of each type:

Standardized test questions	Performance assessment problems
How much change will you get if you have $6.55 and spend $4.32? () $2.23 () $2.43 () $3.23 () $10.87	Suppose you couldn't remember what 8 x 7 is. How could you figure it out?
A good title for this story would be-- () The Pet Store () A Strange Occupation () Going on a Trip () Danny's Pet Snake	At the end of the story, the little old man and the little old woman go on living in the little old house. Do you think this is a good ending? Tell why or why not.

Parents were then asked to rate their approval or disapproval of each type of measure. Respondents were not forced to choose one type of measure over the other. In fact, as can be seen from the data in Table 8, many parents liked both. (As a consequence, oblique factor analyses showed no correlation between the "standardized testing" factor and the "performance assessment" factor.) Although the majority of parents approved of both types of measures, performance assessments had higher approval ratings than did standardized tests, especially in the strongly approve category. For example, in mathematics 18% strongly approved of standardized tests whereas 31% strongly approved of the use of performance assessments.

To try to get at the issue of the influence of assessment formats on instruction, parents were also presented with the following choices:

> Currently in the nation there is some concern that standardized tests have had an undue influence on what gets taught in school, because of too much "teaching to the test." **IF** it is true that tests influence instruction, indicate below whether you prefer for children to be taught mathematics and reading using questions like Box A or Box B:

Box A contained the sample standardized test questions for each subject and Box B contained the performance assessment problems.

As shown in Table 8, more parents preferred the use of questions like those on performance assessments for instructional purposes than preferred the use standardized-test type questions. Although the preference for more open-ended questions is quite large in reading (58% favored performance assessments compared to 21% liking standardized tests), the difference was narrower for instruction in math (44% favored performance assessments over 31% for standardized tests). The tendency for a subgroup of parents to differentiate their choices by subject area is consistent with a pattern found later in the interview data and can clearly be connected in more extended interview responses to parents' conception of mathematics as a one-right-answer subject area. A different, small subgroup of parents resisted choosing between the two types and added comments in the margin emphasizing that a combination of both should be used.

Table 8

Parent Questionnaire Approval Ratings of Standardized Tests and Performance Assessments (n=105)

Suppose the school district were trying to decide whether to use standardized tests, performance assessments, or both standardized tests and performance assessments in all its elementary schools. Indicate below the degree of your approval or disapproval of each type of test or assessment.

Do you approve or disapprove of:	Strongly disapprove	Disapprove	Neutral	Approve	Strongly approve
12. using standardized tests with questions like those in **Box A** to measure your child's achievement in math?	2%	10%	24%	46%	18%
13. using performance assessments with questions like those in **Box B** to measure your child's achievement in math?	6%	5%	20%	38%	31%

Do you approve or disapprove of:	Strongly disapprove	Disapprove	Neutral	Approve	Strongly approve
14. using standardized tests with questions like those in **Box A** to measure your child's achievement in reading?	2%	11%	22%	47%	18%
15. using performance assessments with questions like those in **Box B** to measure your child's achievement in reading?	2%	7%	15%	47%	30%

Currently in the nation there is some concern that standardized tests have had an undue influence on what gets taught in school, because of too much "teaching to the test." IF it is true that tests influence instruction, indicate below whether you would prefer for children to be taught mathematics and reading using questions like Box A or Box B:

	Strongly prefer instruction like Box A	Prefer instruction like Box A	Neutral	Prefer instruction like Box B	Strongly prefer instruction like Box B	Blank/ Missing
16. For math instruction	15%	16%	21%	27%	17%	4%
17. For reading instruction	9%	12%	17%	37%	21%	4%

Once again we wished to examine whether the lowered enthusiasm for standardized tests in this sample compared to the nation could explain the more positive response to performance assessments. Therefore, responses to the above questionnaire items were again considered separately for the subgroups that "Favor" and "Oppose" standardized national tests on the Gallup Poll. Not surprisingly, the group that favors standardized tests gave higher approval ratings to standardized test questions than did the Oppose group (Math: = $\bar{X}$ 4.10, 2.97 respectively; Reading: $\bar{X}$ = 4.13, 2.87 respectively). However, as suggested above by zero correlations between the two factors, approval of standardized tests does not imply disapproval of performance assessments. Parents who favor standardized national tests rated performance assessments just as highly as did parents who oppose standardized national tests (Math: $\bar{X}$ = 3.92, 3.61 respectively; Reading: $\bar{X}$ = 3.90, = 3.87 respectively). Yet, when forced to choose between the two types of questions for use in classroom instruction, there were significant differences between the two groups. Parents who favor standardized tests rated the two types of measures about equally in the context of math instruction ($\bar{X}$ = 2.72) whereas parents who oppose standardize national tests expressed a stronger preference for the use of performance assessments in mathematics instruction ($\bar{X}$ = 3.80). In reading, the pattern was similar ($\bar{X}$ = 3.17, 3.93 respectively). Thus population shifts in the proportion endorsing standardized testing do affect the degree of approval of performance assessments in some contexts, but these analyses suggest that even if 100% of the national population answered yes to the need for standardized national tests, that answer would signify approximately equal approval of performance assessments, not disapproval of them.

Parent Reactions to Standardized Tests and Performance Assessments: Interview Results

It was the impression of the interviewers that almost all parents were intrigued by the opportunity as part of the interview to have a close look at both standardized test questions and performance assessments for third graders. Despite being presented with these examples near the end of the interview, most parents took time to look through the materials carefully. Many worked through the sample problems and asked questions about how they were administered typically or whether this was third-grade work. As can be seen

from sample transcript data in Table 9, parents often pointed to the two different types of questions or used specific problems as examples when explaining their preferences for standardized tests or performance assessments.

Because the purpose of the interview was to understand the reasoning behind parent preferences, qualitative analysis was used to develop categories representing different positions. Entire transcript segments from this portion of the interview were read and distilled so that shortened segments could be used that still contained all key ideas from the longer transcripts. Each representative segment was given a shortened label; similar segments were sorted into categories resulting in the final categorization scheme shown in Table 9. In addition, a separate coding system was developed for key ideas because it was discovered that many features of the two types of tests, cited repeatedly by parents in their discussions, bore little relation to parents' final recommendations or preferences about which type of measure should be used. For example, parents in both the standardized testing and performance assessment camps commented on the worry that performance assessments might be unfair because of the extra demand they placed on students to write their answers.

Table 9

Parent Interview Evaluations of Standardized Test Questions and Performance Assessments: Categories of Responses (n=33)

Strongly prefers standardized tests (n = 2)
Prefers standardized tests (Likes both) (n = 1)
Standardized tests Math/ Performance assessment Reading (n = 4)
Standardized tests Reading/ Performance assessment Math (n = 1)
Both: Likes both, Wants both in instruction (n = 2)
Prefers performance assessments (Likes both) (n = 11)
Strongly prefers performance assessments (n = 11)

The categories in Table 9 are ordered roughly from the position most favorable toward standardized tests to that most favorable toward performance assessments. Only 3 of the 33 parents or parent dyads preferred the use of standardized tests for both district and instructional purposes. Respondents in this category saw standardized tests as more cut and dried, more aligned to instruction, and easier (because having the answers there made it clear what was expected). Another subgroup of parents preferred standardized tests for math and performance assessments for reading. This group of parents liked the opportunity for children to express themselves and the possibility for more than one right answer for reading assessment but maintained that there is one right answer for math. Two sample transcript segments are shown for this category in Table 9. Like them, a third respondent explained: "In mathematics, I think this type [ST] is probably the best . . . because math is pretty basic as far as having the right answer and you have to have the right answer. With this [PA reading] they can use their imagination and they can tell you a story the way they see it. It doesn't always have to be one way." One respondent preferred standardized tests for reading and performance assessments for math. Two responses were assigned in the "middle" category, liking both types of measures for both district purposes and for classroom instruction.

By far the majority of respondents preferred performance assessments. Twelve interview segments were placed in the category "Prefers Performance Assessments (Likes Both)" and another eleven responses were in the "Strongly Prefers Performance Assessments" category. Although the "Prefers PA (Likes Both)" category was heterogeneous, the dominant response was to approve of both kinds of measures being used for district purposes but to prefer that performance assessments be used for classroom instruction. Across responses in all categories the most frequently mentioned feature of performance assessments is that they make children think. As shown by the counts for each code in Table 10, this was noted by 10 of the 12 responses in the "Prefers PA" category. In addition to the several responses shown in Table 9, other respondents in this category explained why they preferred performance assessments for instruction:

Table 10

Key Features of Standardized Tests and Performance Assessments Identified in Interviews by Subgroups of Parents

Codes	Strongly prefers ST (n=2)	Prefers ST (Likes both) (n=1)	ST Math/ PA Reading (n=4)	ST Reading/ PA Math (n=1)	Likes both (n=2)	Prefers PA (Likes both) (n=12)	Strongly prefers PA (n=11)
Guess				1		6	9
Y/N	2	1	3			5	2
Easy	1		1	1	1	2	2
Obj	1					2	2
Support		1				3	1
Know (ST)					1	1	1
Norm						2	1
Think		1	2	1	2	10	8
Imagin	1		4			5	4
Diag	1		1			8	6
Hard	1	1		1	1	5	1
Unfair	1		1		1	5	2
Subj	1				1	3	3
Know (PA)					1	4	4
Align		1	1			4	3

Note. The following codes were used in the transcript analysis for frequently mentioned characteristics of standardized tests and performance assessments:

Guess	ST allows students to guess
Y/N	ST have black and white, yes-no answers
Easy	ST are easy
Obj	ST are objective, have a clear right answer
Support	Having the answers on ST helps children know what's expected
Know (ST)	ST shows that children have skills, can comprehend
Norm	ST provides national comparison
Think	PA makes students think
Imagin	PA lets children express themselves, show imagination, be creative
Diag	PA shows you where children need help
Hard	PA are difficult, + a challenge, – too hard
Unfair	PA reading and writing demands might be unfair to some children
Subj	PA are difficult to grade, subjective
Know (PA)	PA shows children really have the concepts because they have to explain
Align	Would have to teach more like (PA) to test that way

> I think this [PA] would make them think more.
>
> I think in order to learn any kind of subject you have to have concepts down and I think number 2 is going to show how to develop the concepts better....You need to get those basics. . . . But I do think this [PA] is going to make them think more.
>
> I just think, like here [ST] they get to choose where here [PA] they really have to think about it and you're really gonna know that they know exactly what they read or what have you.

Even respondents who preferred standardized tests for other reasons often noted that performance assessments would stimulate children's imagination or make them have to think.

The substance of comments regarding both standardized tests and performance assessments was similar for parents in both the "Prefers PA" and "Strongly Prefers PA" categories with the distinction tending to be that parents in the first group gave approval ratings to standardized tests for district purposes. Both groups, however, criticized standardized tests because it was possible to guess and because the results of such tests would not tell you much about the process of a student's learning or where they were failing to comprehend. In contrast, we used the code "Diag" to note the many instances where parents commented on the diagnostic value of performance assessments. For example, from one of the "Prefers PA" responses not shown in Table 9:

> The other tests [PA math] kind of makes them tell you the concept, not just the right answer. I like the "explain your choice," or "what would you tell Adam type" questions. . . . this would give a teacher more information to think about, especially on the concepts that they haven't quite grasped yet.

Several parents also commented that standardized tests were more objective in contrast to performance tests that would be either difficult to grade fairly or more time consuming to grade. In some cases, however, parents went on to say that this was exactly the kind of attention or commitment that was needed in education. Concerns about grading were voiced along with the concern that reading level might create problems for some students on the math assessment or that writing might be a problem on both performance assessments. Although these concerns were mentioned as problems to be resolved, for the most part they did not appear to affect parents' enthusiasm for

using performance assessments in instruction. For some parents in the "Prefers PA (Likes Both)" category, however, these features were cited specifically as the reason that both types of tests should be used at least for district purposes. "Different children learn in different ways." Some children, especially those who "are not good with words," would be helped by having the answers there so they could show that they understood. (Answers of this type were coded "Support" in Table 10.)

Conclusions

The purpose of the study was to examine parent opinions about standardized tests and new performance assessments in greater depth than can be understood from national survey data. The classic Gallup Poll question showing a high percentage of citizens and public school parents in favor of standardized national tests is often interpreted as a mandate for external, machine-scorable accountability measures. What was discovered in this study is that parents' favorable ratings of standardized national tests do not imply a preference for such measures over other less formal sources of information for monitoring their child's academic progress *or* for judging the quality of education provided at their local school. Approval of standardized tests likewise does not imply disapproval of performance assessments.

In this study, third-grade parents considered report cards, hearing from the teacher about their child's progress, and seeing graded samples of student work to be much more useful in learning about their child's progress than standardized tests. In interview data, parents often mentioned the need for comparative information to know how to interpret their own child's progress, but they trusted the teacher to tell them how their child was doing in relation to grade-level expectations or to other children in the class. Comparison to external or national norms was mentioned rarely. Even for accountability purposes, the usefulness ratings for standardized tests increased but did not equal parents' high ratings of talking to the teacher and seeing student work. According to parents, these sources of information are indicators of school quality because they allow them to see what is being taught and what expectations are set by the classroom teacher.

Because preliminary analyses of Gallup Poll results showed the third-grade parent sample in this study to be less favorably disposed toward

standardized tests than the national sample, follow-up analyses were conducted to determine whether high parent ratings of informal sources of information could be attributed to the particular nature of this sample. However, even the 46% subsample of parents who favored standardized national tests on the Gallup Poll rated report cards, talking to the teacher, and seeing student work as more useful that standardized tests.

Using both questionnaire and interview formats, different samples of parents were provided with specific examples of the types of questions used on standardized tests and on performance assessments in both reading and mathematics. While a majority of parents approve of both standardized tests and performance assessments, approval ratings were stronger for performance assessments. Again this overall pattern could not be explained as merely a selection bias in this particular sample. The group that favored national tests rated the two types of measures equally. A pervasive theme in the interview data was that performance assessment problems "make children think" and are likely to give teachers better insights about what children are understanding and where they are struggling. Parents commented frequently about the desirability of having children explain their answers in mathematics and being encouraged to express themselves in response to stories they read. Standardized tests were seen as easier and more supportive by some parents because having answer choices communicates what's expected and allows children who aren't very verbal to show what they know; at the same time, parents complained frequently that multiple-choice questions allow children to guess the right answer "25% of the time."

When parents in this study had a chance to look closely at performance assessment problems, most endorsed their use for district purposes and especially preferred their use in classroom contexts. Therefore, survey data showing approval of standardized national tests should not be taken to mean that parents are opposed to the use of alternative measures. A few parents expressed concern about the subjectivity of performance measures but lack of rigor—sometimes evoked by the term "alternative assessment"—was never mentioned. Instead, when parents looked at specific performance assessment problems, they were more likely to comment that the questions seemed hard or challenging. It was the impression of the interview team, in fact, that seeing real examples of both types of tests affected parent responses. The

questionnaire and interview demonstrations served an educative function. Parents might have answered differently if they had to rely on their own ideas about what performance assessments look like.

References

Borko, H., Flory, M., & Cumbo, K. (1993, April). *Teachers' ideas and practices about assessment and instruction.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Darling-Hammond, L. (1988). Accountability and teacher professionalism. *American Educator, 12,* 8-13, 38-43.

Elam, S. M., Rose, L. C., & Gallup, A. M. (1992). The 24th annual Gallup-Phi Delta Kappan Poll of the public's attitudes toward the public schools. *Phi Delta Kappan,* (September), 41-53.

Flexer, R.J., & Gerstner, E.A. (1993, April). Dilemmas and issues for teachers developing performance assessments in mathematics. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Hiebert, E.H., & Davinroy, K. (1993, April). *Dilemmas and issues in implementing classroom-based assessments for literacy.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Madaus, G. S., West, M. M., Harmon, M. C., Lomax, R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4-12: Executive summary.* Boston: Boston College, Center for the Study of Testing.

McNeil, L. M. (1988). Contradictions of control, Part 3: Contradictions of reform. *Phi Delta Kappan, 69,* 478-485.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction.* Boston: Kluwer Academic Publishers.

Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 73,* 232-238.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46,* 41-46.